目　　录

第一部分　机理分析法

第二部分 数据分析法

第三部分 仿真与其他方法

第一章　数学建模的方法论基础

随着科学技术对研究对象的日益精确化、定量化和数学化，随着电子计算技术的广泛应用，数学模型已成为处理科技领域中各种实际问题的重要工具，并在自然科学、工程技术科学与社会科学的各个领域中得到广泛应用. 诸如经济、管理、工农业，甚至社会学等. 什么是"数学模型"，如何建立数学模型，是科技工作者极感兴趣的问题.

§1.1　数学模型

数学模型，就是针对或参照某种问题（事件或系统）的特征和数量相依关系，采用形式化语言，概括或近似地表达出来的一种数学结构.

数学模型因问题不同而异，建立数学模型也没有固定的格式和标准，甚至对同一个问题，从不同角度、不同要求出发，可以建立起不同的数学模型. 因此与其说数学建模是一门技术，不如说是一门艺术. 它需要熟练的数学技巧、丰富的想象力和敏锐的洞察力，需要大量阅读、思考别人做的模型，尤其要自己动手，亲身体验.

建立数学模型一般有如下要求：

1° 足够的精度，即要求把本质的关系和规律反映进去，把非本质的去掉.

2° 简单、便于处理.

3° 依据要充分，即要依据科学规律，经济规律来建立公式和图表.

4° 尽量借鉴标准形式.

5° 模型所表示的系统要能操纵和控制，便于检验和修改.

用数学方法研究实际问题，需要对这些问题进行识别和考虑最适合的或比较好的提法，这不仅需要相应领域的数学理论和方法，也需要相应领域的专业知识.因此建立模型的工作，常常是由数学家与有关专家共同完成.

建立数学模型的一般步骤是：

第一步　对问题（事件或系统）进行观察，想象其运动变化情况，用非形式语言（自然语言）进行描述，初步确定描述问题的变量及相互关系.

第二步　确定问题的所属系统（力学系统、生态系统、管理系统等），模型大概的类型（离散模型、连续模型、随机模型等）以及描述这类系统所用的数学工具（图论方法、常微分方程等），提出假说.

第三步　将假说进行扩充和形式化，选择具有关键性作用的变量及其相互关系（主要矛盾），进行简化和抽象，将问题的内在规律用数字、图表、公式、符号表示出来，经过数学上的推导和分析，得到定量（或定性）关系，初步形成数学模型.

第四步　根据现场试验和对试验数据的统计分析估计模型参数.

第五步　检验修改模型.这是在反映问题的真实性与便于数学处理之间的折衷过程.模型只有在被检验、评价、确认基本符合要求后，才能被接受；否则需要修改模型，这种修改有时是局部的，有时甚至要推倒重来.

建立数学模型，可能会涉及到许多数学分支.一个问题，往往可以利用不同方法建立不同的模型.因此绝对的分类，对于建立数学模型是不利的.但是大致的分类，对初学者，在确立原型所属系统和采用数学工具时，会有一定的帮助.数学模型有多种分类方法：

按时间变化对模型的影响，可分为时变与时不变模型，静态与动态模型等.

按变量情况可分为离散模型与连续模型，确定性和随机性模型等.

按实际系统与周围环境相互关系可分为自治的和非自治模型.

按研究方法和对象的数学特征，可分为优化模型、逻辑模型、稳定性模型、扩散模型等.

按研究对象的实际领域可分为人口模型、交通模型、生态模型、经济模型、社会模型等.

模型的修改与化简，是建模中技巧性较强的环节. 由于实际情况是复杂多变的，往往不能简单套用现有模型. 例如，有的参数在某个场合容易得到，而在另一场合却得不到，这就迫使人们改用其他形式的模型；有时在构造模型的过程中发现必须拥有这样或那样的数据，或指出模型应朝哪一个方向修正；有时，虽然复杂的模型已经构出，但作试验或求解却十分困难，这也迫使人们采用较简单的近似模型.

常用简化模型的方法有：

1° 除去一些变量

在机理分析中，在一定条件下，常将描述分布参数系统的偏微分方程，简化为集中参数的常微分方程.

在统计分析中，则采用主成分分析法、向后回归法(淘汰法)和逐步回归方法[2][18]，以减少变量个数. 或在建模之前，采用正交试验方法，在众多因素(变量)中找出对指标有显著影响的少量因素再进行优选试验，进而建立模型.

2° 合并一些变量

在构造模型时，把一些性质相同或相似的变量合并成少数有代表性的变量. 尽管这样做降低了模型的精度，但只要能满足建模的基本要求，则是可行的. 例如在经济系统建模中，经过多年研究探索，将国民经济上千个部门合并成 61 个变量.

3° 改变变量的性质

常用的方法是，把某些非主要的或暂时的变量看作常量，把连续变量看作离散变量，或把离散变量看作连续变量.

4° 改变变量之间的函数关系

当处理非线性问题遇到困难时，或建模精度要求不高时，常将非线性函数在某一点处展开(Taylor 展开)，取前两项作为近似表达式，即用线性关系逼近非线性关系式. 这一线性化方法在工程界被广泛采用. 也可以采用二次函数或其他研究比较透彻的函数逼近，而使模型简化.

在随机性模型中，常采用一些熟悉的概率分布函数，如正态分布、指数分布等去代替不太好处理的概率分布函数.

5° 改变约束关系

为简化模型有时还可以对变量的约束条件加以改变，如增加一些约束，或去掉一些约束，对约束进行一些修改等等. 例如在求解数学规划问题时，若要求目标函数的极大值，而真正解不一定能找到时，则增加约束后求得的可行解一般是偏低的，称之为保守解或悲观解. 去掉一些约束求得的解往往偏高，称之为冒进解或乐观解. 虽然它们都不是问题的真正的解，但可以通过他们来了解真正解的范围，这对问题进行初步评价是有用的.

6° 模型结构的转换

若某种模型在理论上很漂亮，但求解很困难，甚至无法求解，或者某种模型，要求具备某种数据，而这种数据不具备或不易得到，我们只有改用其他形式的模型，即改变模型的结构.

模型结构的转换，需要在对问题透彻理解和想象的基础上，实现视角的转换，即从不同的角度观察问题，进而采用不同的数学工具来描述同一问题.

在建模时，能否用数学工具描述某一问题的特征是建模的前提. 当根据观测数据对回归模型的参数或时序模型的参数进行估计时，系统可辨识性问题也就同时提出来了. 当根据某物理场的信息估计相应偏微分方程中的某些系数时，如场的存在范围或边界

条件，我们也遇到了数学物理反问题的适定性问题，…．这些数学建模的理论问题留待在专门的问题中研究．这里仅围绕数学建模的方法展开讨论．

§1.2 建模的逻辑思维方法

从对数学模型的要求、建模的过程与步骤来看，要建立数学模型，应具备下述五个方面的能力：

1° 分析综合能力；

2° 抽象概括能力；

3° 想象洞察能力；

4° 运用数学工具的能力；

5° 通过实践验证数学模型的能力．

建立数学模型是一种积极的思维活动，从认识论角度看，是一种极为复杂且应变能力很强的心理现象，因此没有统一的模式，没有固定的方法，其中既有逻辑思维，又有非逻辑思维．建模过程大体都要经过分析与综合、抽象与概括、比较与类比、系统化与具体化的阶段，其中分析与综合是基础，抽象与概括是关键．从逻辑思维来说，抽象、归纳、演绎、类比等形式逻辑的思维方法大量被采用．熟悉这些基本方法，无疑对提高建模能力会有帮助．下面试图以一些实例说明这些方法的应用．当然这些实例本身是多种方法的结果，并不能绝然划分到某一方法类中．

一、抽象

科学研究就是要揭示事物的共性和联系的规律，因此就要忽略每个具体事物的特殊性，着眼于整体和一般规律．

例 1-1 人们在日常生活中，经常会遇到这样一个问题：有四条腿的家具，如椅子、桌子等，往往不能一次放稳，只能有三只脚着地，需要旋转调整几次，方可以使四只脚着地，放稳．这个看来似乎

与数学无关的现象能用数学语言表述，并用数学工具证实吗？

数学建模的关键是用数学语言把四只脚同时着地的条件和结论表示出来.

1° 椅子的位置和调整的表述. 注意到椅子脚连线成正方形，以中心点为对称点，正方形绕中心的旋转表示了椅子位置的改变（可假设椅子位置调整中只有旋转而没有平移，因为在实际问题中只要旋转调整便可放稳）. 因此可以用旋转角度这一变量表示椅子的位置. 在图 1-1 中，$ABCD$ 为椅子初始位置，$A'B'C'D'$ 为椅子绕中心点 o 旋转 θ 角后的位置.

2° 椅脚着地的数学表示. 显然若用变量表示椅脚与地面的距离，当此变量为零时，就表示椅脚着地. 这样需引进四个变量，且均为 θ 的函数（因为椅子的位置不同时，椅脚与地面的距离不同）.

现在考虑化简. 由于正方形是中心对称的，只要假定两个距离函数即可. 设 A、C 两脚与地面距离之和为 $f(\theta)$，B、D 两脚与地面距离之和为 $g(\theta)$，显然 $f(\theta), g(\theta) \geqslant 0$.

对三只脚着地和四只脚着地的描述. 由于椅子在任何位置至少有三只脚着地，所以对于任意的 θ，$f(\theta)$ 和 $g(\theta)$ 中至少有一个为零，因此恒有 $f(\theta) \cdot g(\theta) = 0$. 当 $\theta = 0$ 时，不妨设 $g(\theta) = 0, f(\theta) > 0$. 若四只脚一样长，则旋转 90°后，只是两对角线互换，因此当 $\theta = \pi/2$ 时，$f(\theta) = 0, g(\theta) > 0$. 在 $\theta = \theta_0$ 四只脚着地时，$f(\theta_0) = g(\theta_0) = 0$.

3° 函数 $f(\theta)$ 与 $g(\theta)$ 的性质. 假设地面高度是连续变化的，则 $f(\theta)$ 和 $g(\theta)$ 为 θ 的连续函数.

将上述分析中的假设和模型整理出来.

1. 模型假设

1° 椅子四条腿一样长（这样椅子在绕中心旋转时，仅与 θ 角有关，而不会因四条腿不一样长，而与椅腿有关），椅脚与地面接触处可视为一个点（只考虑几何位置），四脚的连线呈四方形.

2° 地面高度是连续变化的，即为连续曲面，沿任何方向都不

会出现间断(保证了 f、g 的连续性).

3° 对于椅脚的间距和椅腿的长度而言,地面是相对平坦的,椅子在任何位置至少有三只脚同时着地.

2. 模型的构成

将用自然语言描述的现象,翻译成形式化的数学语言.

令 $f(\theta)$ 为 A、C 与地面距离之和,$g(\theta)$ 为 B、D 与地面距离之和. f、g 是 θ 的连续函数,则问题表述为:

已知连续函数 $f(\theta)$ 和 $g(\theta)$,$\theta\in[0,\pi/2]$,满足

$$f(\theta)\cdot g(\theta)=0 \quad \forall\ \theta\in[0,\pi/2]$$

且 $f(0)\geqslant 0, g(0)=0;\quad f(\pi/2)=0, g(\pi/2)\geqslant 0$

求证:存在 $\theta_0\in[0,\pi/2]$,使 $f(\theta_0)=g(\theta_0)=0$.

3. 模型求解

现在问题的数学求解也清楚了.

令 $h(\theta)=f(\theta)-g(\theta)$,则 $h(0)>0$,而 $h(\pi/2)<0$,由于 f、g 是连续的,故 h 亦为连续. 根据连续函数的中值定理知,必存在 θ_0,$0<\theta_0<\frac{\pi}{2}$,使 $h(\theta_0)=0$,即 $f(\theta_0)=g(\theta_0)$. 又因为 $f(\theta_0)\cdot g(\theta_0)=0$,故

$$f(\theta_0)=g(\theta_0)=0.$$

问题得到解决. 在该问题的建模中巧妙的是用一元变量 θ 表示椅子的位置,以及用两个函数表示椅子四脚与地面的距离. 根据实际经验,椅子或桌子亦可以是长方形,由此可看出利用正方形的中心对称性及旋转 90°不是本质性的.

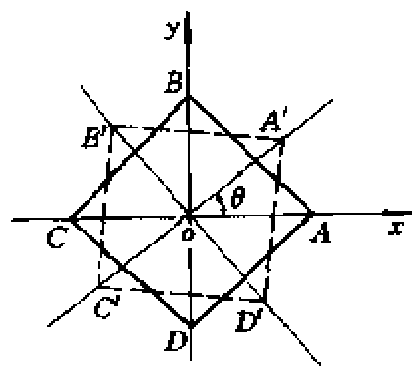


图 1-1 椅子的放置

二、归纳

就人类总的认识秩序而言,总是先认识某些特殊现象,然后过

渡到对一般现象的认识.归纳就是从特殊的具体的认识推进到一般的抽象的认识的一种思维方式，它是科学发现的一种常用的有效的思维方式.归纳的前提是存在单个的事实或特殊的情况，所以归纳是立足于观察、经验或实验的基础上的.另外，归纳是依据若干已知的不完全的现象推断尚属未知的现象，因此结论具有猜测的性质，然而它却超越了前提包含的内容.

开普勒第三定律的发现，可视为归纳法的典型例子.

例 1-2 开普勒第三定律的发现.

第谷·布拉赫(1546—1601)观测行星运动，积累了 20 年的资料.开普勒(1571—1630)作为他的助手，运用数学工具分析研究这些资料，发现火星的位置与根据哥白尼的"行星绕太阳的运行轨道是圆形的"理论所计算的位置相差 8 弧分.在深入分析的基础上，他于 1609 年归纳出所谓开普勒第一定律：各行星分别在不同的椭圆轨道上绕太阳运行.太阳位于这些椭圆的一个焦点上.以及开普勒第二定律：单位时间内，太阳——行星行径扫过的面积是常数(对一颗行星而言).为了寻求行星运动周期与轨道尺寸的关系，他将当时已发现的六大行星的运行周期和椭圆轨道的长半轴列成表格，如表 1-1 所示.经反复研究.终于总结出第三定律：行星运行周期的平方与其椭圆轨道长半轴的三次方成正比.

表 1-1　六大行星运行周期和椭圆轨道的长半轴

行星	周期 T	长半轴 a	T^2	a^3
水星	0.241	0.387	0.058	0.058
金星	0.615	0.723	0.378	0.378
地球	1.000	1.000	1.000	1.000
火星	1.881	1.524	3.54	3.54
木星	11.862	5.203	140.7	140.85
土星	29.457	9.539	867.7	867.98

显然开普勒在总结上述规律时使用的是不完全归纳法，在理论证明后才成为定律，但归纳所得到的猜测，却具有科学发现的重

大意义. 在计算机和计算方法迅速发展的今天，这种归纳也可以用一种数学方法得到，这就是数据分析法中的非线性回归，这一点将在第八章中讨论.

三、演绎

演绎推理是由一般性的命题推出特殊命题的推理方法. 演绎推理的作用在于把特殊情况明晰化，把蕴涵的性质揭露出来，有助于科学的理论化和体系化.

牛顿以微积分为工具，在开普勒三定律和牛顿力学第二定律的基础上，演绎出万有引力定律，这一定律成功地定量地解释了许多自然现象，也为其后一系列的观测和实验数据所证实.

例 1-3 万有引力定律的产生.

牛顿认为一切运动都有其力学原因，开普勒三定律的背后必定有某个力学规律起作用，他要构造一个模型加以解释.

以为原点建立极坐标系，向径 r 表示位置，如图 1-2 所示.

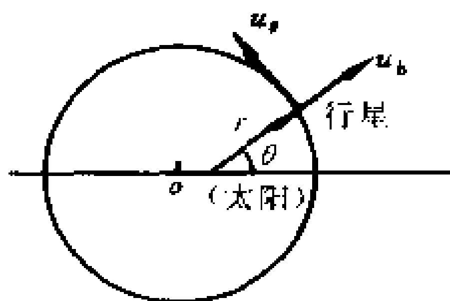


图 1-2　六系中的行星运道

将开普勒三定律作为假设 Ⅰ、Ⅱ、Ⅲ，牛顿力学第二定律作为假设Ⅳ，它们可表示为

Ⅰ）轨道方程为

$$r=\frac{p}{1+e\cos\theta} \tag{1}$$

其中，$p=\frac{b^2}{a}$，$b^2=a^2(1-e^2)$，a 为长半轴，b 为短半轴，e 为离心率.

Ⅱ）
$$\frac{1}{2}r^2\theta=A \tag{2}$$

其中 A 是单位时间内向径 $\boldsymbol{r}$ 扫过的面积，对某一颗行星而言，A 是常数. $\dot{\theta}$ 表示 θ 对时间 t 的导数.

Ⅲ） $$T^2=ka^3 \tag{3}$$

其中 T 是行星运行周期，k 是绝对常数.

Ⅳ） $$\boldsymbol{f}\propto\ddot{\boldsymbol{r}} \tag{4}$$

表示太阳和行星间的作用力 $\boldsymbol{f}$ 与加速度 $\ddot{\boldsymbol{r}}$ 的方向一致，与 $\ddot{\boldsymbol{r}}$ 的大小成正比. 现在试图从这四条假设出发，寻找太阳与行星间作用力的方向和大小应满足的关系：即 $\ddot{\boldsymbol{r}}$ 的关系式.

选取基向量

$$\begin{cases}\boldsymbol{u}_r=\cos\theta\boldsymbol{i}+\sin\theta\boldsymbol{j}\\ \boldsymbol{u}_\theta=-\sin\theta\boldsymbol{i}+\cos\theta\boldsymbol{j}\end{cases} \tag{5}$$

如图 1-2 所示，于是 $\boldsymbol{r}=r\boldsymbol{u}_r$ （6）

因为

$$\dot{\boldsymbol{u}}_r=-\sin\theta\cdot\dot{\theta}\boldsymbol{i}+\cos\theta\cdot\dot{\theta}\boldsymbol{j}=\dot{\theta}\boldsymbol{u}_\theta \tag{7}$$

$$\dot{\boldsymbol{u}}_\theta=-\cos\theta\cdot\dot{\theta}\boldsymbol{i}-\sin\theta\cdot\dot{\theta}\boldsymbol{j}=-\dot{\theta}\boldsymbol{u}_r \tag{8}$$

由(6)(7)式得到行星运动的速度和加速度

$$\dot{\boldsymbol{r}}=\dot{r}\boldsymbol{u}_r+r\dot{\theta}\boldsymbol{u}_\theta \tag{9}$$

$$\ddot{\boldsymbol{r}}=(\ddot{r}-r\dot{\theta}^2)\boldsymbol{u}_r+(r\ddot{\theta}+2\dot{r}\dot{\theta})\boldsymbol{u}_\theta \tag{10}$$

据(2)式，有

$$\dot{\theta}=\frac{2A}{r^2}, \tag{11}$$

$$\ddot{\theta}=\frac{-4A\dot{r}}{r^3} \tag{12}$$

由(11)和(12)式知(10)式右端第二项 $r\ddot{\theta}+2\dot{r}\dot{\theta}=0$，故有

$$\ddot{\boldsymbol{r}}=(\ddot{r}-r\dot{\theta}^2)\boldsymbol{u}_r$$

据(1)和(2)式，可得

$$\dot{r}=\frac{r^2}{p}e\sin\theta\cdot\dot{\theta}=\frac{2A}{p}e\sin\theta \tag{14}$$

$$\ddot{r}=\frac{2A}{p}e\cos\theta\cdot\dot{\theta}=\frac{4A^2}{r^3}\left(1-\frac{r}{p}\right) \tag{15}$$

将(11)(15)式代入(13)式，得

$$\ddot{\boldsymbol{r}}=-\frac{4A^2}{pr^2}\boldsymbol{u}_r \tag{16}$$

将(16)式与(5)(7)式相比较知，太阳对行星的作用力 f 的方向与向径 r 方向正好相反，即 f 在太阳与行星的连线方向上，指向太阳；f 的大小与太阳——行星间距离的平方成反比.

下面进一步证明(16)式中的比例系数 A^2/p 是绝对常数（A 和 p 都不是绝对常数，其数值取决于所讨论的是哪一颗行星）.

根据 A 和(2)式中 a、b 的定义，任一行星的运行周期 T 满足

$$TA=\pi ab \tag{17}$$

由(2)、(4)式和(17)式可得

$$\frac{A^2}{p}=\frac{\pi^2a^2b^2}{T^2p}=\frac{\pi^2a^2b^2}{ka^3}\cdot\frac{a}{b^2}=\frac{\pi^2}{k}$$

式中 π 和 k 皆为绝对常数，这说明引力的比例系数对"万物"是同一常数. 万有引力定律得到证明.

四、类比

类比是在两类不同的事物之间进行对比，找出若干相同或相似点之后，推测在其他方面也可能存在相同或相似之处的一种思维方式. 由于类比是从人们已经掌握了的事物的属性，推测被正在研究中的事物的属性，所以类比的结果是猜测性的，不一定可靠，但它却具有发现的功能，是创造性思维的重要方法.

例 1-4　电话系统模型

本世纪初，由于自动电话的出现，通话需求与通话服务供给的平衡问题，成了研究的热门. 丹麦数学家埃尔朗(A. K. Erlang)，在物理学家吉布斯(Gibbs)统计平衡概念的启发下，超越一般组合分析计算的方法，运用类比的思维，把统计平衡概念借用过来，建立了呼叫生灭过程的模型.

埃尔朗把封闭系统热分子的渗透扩散比作电话系统呼叫的生灭. 一个呼叫的发生，好像一个分子从液体扩散进入封闭系统中的气体，而一个呼叫的结束，又好像一个热分子从气体中渗透到液体中，这样他就把热力学统计平衡模型全部搬过来，建立起电话呼叫

统计平衡的方程. 如图 1-3 所示就是一个类比对照.

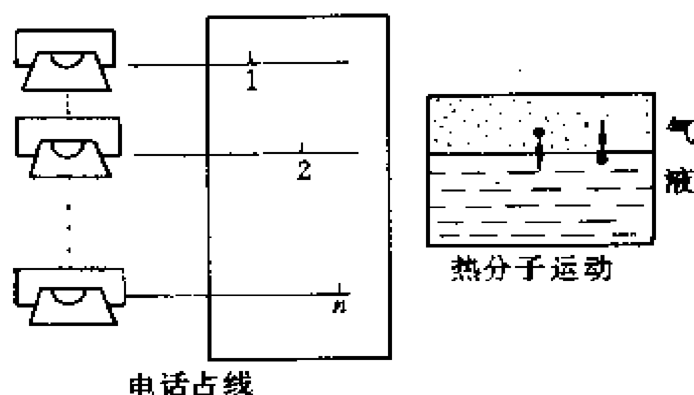


图 1-3　电话占线与热分子运动的类比

现在考虑单位时间内呼叫需求,即单位时间内到达的呼叫数,用 λ 表示. 对于电话服务,可以用单位时间内通话完毕的通话次数 μ 表示,并假定两者使用同一单位时间度量. 这正像在封闭容器内汽化的热分子的汽化率与冷凝热分子的冷凝率一样. 把电话系统内瞬时的正在通话的线路数 n 叫做状态(类似于封闭系统内气体的热分子数). 所谓统计平衡就是由状态 n 到状态 $n+1$ 的概率和由状态 $n+1$ 到状态 n 的概率相等.

令 P_n 表示在 Δt 内系统处于状态 n 的概率,则根据统计平衡,有

$$\lambda \cdot \Delta t \cdot P_n = (n+1)\mu \Delta t P_{n+1}$$

(严格来说,P_n 与 P_{n+1} 与时间 t 有关,但在系统平衡时,即 $\Delta t \to 0$ 时,可以表示平稳值),于是

$$\lambda P_n = (n+1) \cdot \mu \cdot P_{n+1}, \quad n = 0, 1, 2, \cdots$$

有递推公式

$$P_{n+1} = \frac{\lambda}{(n+1)\mu} P_n = \frac{(\lambda/\mu)^2}{(n+1) \cdot n} P_{n-1} = \cdots = \frac{(\lambda/\mu)^{n+1}}{(n+1)!} P_0$$

如果系统无限制,则

$$\sum_{n=0}^{\infty} P_n = \sum_{n=0}^{\infty} \frac{(\lambda/\mu)^n}{n!} P_0 = 1$$

故 $$P_0=1\Big/\sum_{n=0}^{\infty}\frac{(\lambda/\mu)^n}{n!}=e^{-\lambda/\mu},$$

$$P_n=\frac{(\lambda/\mu)^n}{n!}e^{-\lambda/\mu}.$$

这是 Poisson 分布.

但系统一般都有限制，$0\leqslant n\leqslant N$，则

$$\sum_{n=0}^{N}\frac{(\lambda/\mu)^n}{n!}P_0=1,$$

$$P_0=\left[1+\lambda/\mu+\frac{(\lambda/\mu)^2}{2!}+\cdots+\frac{(\lambda/\mu)^N}{N!}\right]^{-1},$$

故 $$P_n=\frac{(\lambda/\mu)^n}{n!}\left[1+\frac{\lambda}{\mu}+\frac{(\lambda/\mu)^2}{2!}+\cdots+\frac{(\lambda/\mu)^N}{N!}\right]^{-1}.$$

这就是著名的埃尔朗电话损失率公式. 许多年来，就是用这一公式来设计运用电话系统的通话线路与通话率的. 有关埃尔朗分布，可参考文献[15].

在自然现象中可以看到许许多多可类比的同态现象，如电路振荡系统与机械振荡系统、单振简谐运动与 L-C 振荡等. 康德在《宇宙发现概论》中说："每当理智缺乏可靠论证的思想时，类比这个方法往往指引我们前进."

例 1-5 人体肌肉的类比模型.

当人体肌肉不受力时，其作用类似于无源机械元件. 若施加一外力(例如提起一重物)使肌肉拉伸，此时肌肉呈现弹性机械的特点，肌肉组织的伸缩运动常常伴随着热量的产生和温度的增高. 这些效应表明在肌肉组织内有某种类似于摩擦机构的作用，使得肌肉运动时一部分机械能作功，而另一部分则变为热能.

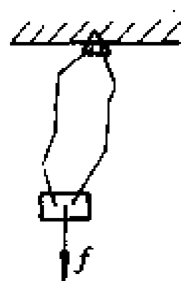
按照以上分析，可以用一个理想的弹簧和一个阻尼器的组合来类比一束肌肉的物理模型，其中弹簧类比于肌肉的弹性，而阻尼器则类比于肌肉的摩擦现象，如图 1-4(a)所示. 其中(c)为力学模型的电路类比模型. 这里利用了机械系统和电路的类比关系：作用力 $f(t)$ 类比于电压源 $e(t)$，阻尼 D 类比于电阻 R，质量 M 类比于

电感 L. 描述图 1-4(b)所示的数学模型是

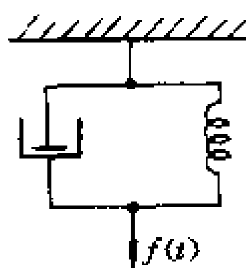
$$f(t)=D\frac{\mathrm{d}y}{\mathrm{d}t}+Ky=Dv+K\int v\mathrm{d}t$$

描述图 1-4(c)所示的数学模型是

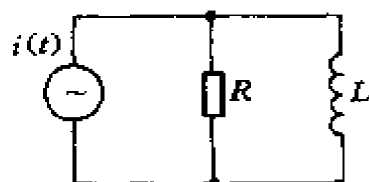
$$i(t)=\frac{1}{R}u(t)+\frac{1}{L}\int u(t)\mathrm{d}t$$



(a)无源肌肉



(b)肌肉的力学类比模型



(c)肌肉的电学类比模型

图 1-4 肌肉的类比模型

五、模拟

模拟是以模型去模拟和仿照原型的合理的结构特性和特殊的功能、原理的一种科学研究方法，它以类比为逻辑基础，或属于类比的一种特殊形式，而在功能、原理的产生或来源的方式上称为模拟.

下面是 Packing 问题的一个启发式(heristic)解法[25].

例 1-6 放置问题.

考虑如下问题：在一个已知的容器中希望能放下 N 个已知不同形状大小的物体，其中界限容器的封闭边境以及各个物体都是不可嵌入的刚性实体. 如果客观上放不下，要求作出放不下的判断，如果客观上放得下，则要求给出每个物体的位置和方向. 这里忽略了 N 个物体之间的内涵，即相互间联系的约束，而认为是独立的，因而可以仅仅研究其几何关系.

这类问题在科学研究与生产实践中具有重要意义，如卫星仪

器舱和坦克内的布置设计以及下料问题等等.但是对于这样一个看来十分清楚的几何问题,在数学上至今还没有一般性的严格的理论解法,这是一个具有 NP 难度的问题(参见第三章§3.4).

我们将放置问题视为一动态过程,即从某初始状态开始的调整过程.这一调整过程可从日常生活中的实例中受到启发.其一是大米的放置.一缸大米,当我们将米缸轻轻摇动,每颗大米都在运动,在调整自己的位置,最终米缸中米的高度降低了,因此又可以多放置适当数量的大米.其二是在一辆拥挤的公共汽车中,在汽车起步、刹车和颠簸过程中,每个乘客都在作微小运动,调整自己的位置,最终大家都感到比开始要自在一些,若能达到彼此都不挤(可以接触),则可认为人都放置好了.由此可将原问题中这 N 个物体视为光滑的弹性实体,将容器想象为充满整个三维空间的光滑弹性物,不过其中因挖去部分实体而形成一个空腔,此空腔的大小形状相同于容器所境界的空间部分.想象这 N 个弹性体挤缩在这个弹性空腔中.如果原始的刚性体放置问题客观上有解,那么这个存在挤压的弹性物体与空腔所构成的体系就会在弹性力的作用下发生一系列的运动,最终有可能使得各个物体与空腔都恢复自己的大小与形状,因而放置问题的定解条件得到满足.

将固结于容器之上的空间笛卡尔坐标系取作绝对坐标系,对每一个物体都按某种方式事先指定一个固结于其上的笛卡尔坐标系,此坐标系的原点选择在物体的几何重心上.第 i 个物体在容器中的状态由以下六个实数所描述:$x_i,y_i,z_i,\theta_i,\varphi_i,\psi_i,i=1,2,\cdots,N$.其中 x_i,y_i,z_i 表示第 i 个物体的几何重心在绝对坐标系之下的坐标;$\theta_i,\varphi_i,\psi_i$ 表示第 i 个物体在绝对坐标系之下的欧拉角,它们表现了此物体所处的方向.因此整个容器体系的一个状态由如下 $6N$ 个实数所描述:$x_i,y_i,z_i,\theta_i,y_i,z_i,i=1,2,\cdots,N$.

我们将物体容器体系的所有状态所构成的集合称作问题的状态空间,对于体系的任意一个确定的状态称为状态空间中的一个点.

引进两物体间的距离的概念. 记 i、j 两物体间的距离为 L_{ij}，其中 $i \neq j, i, j = 1, 2, \cdots, N$，当 i、j 两物体的交的 Lebesgue 测度大于零时，定义 $L_{ij} < 0$，其绝对值等于 i、j 两物体沿几何重心的连线方向以平移方式互相远离直到它们的交的 Lebesgue 测度为零时所经过的长度. 当 i、j 两物体的交的 Lebesgue 测度为零时，我们定义 $L_{ij} \geqslant 0$，其数值等于 i、j 两物体沿其几何重心连线方向以平移方式互相接近直到它们的交的 Lebesgue 测度为 0^+ 时所经过的长度. 如图 1-5 所示.

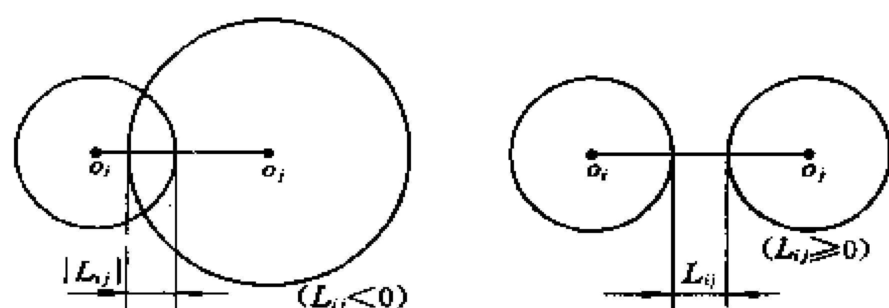


图 1-5　物体间的交的 Lebesque 测度

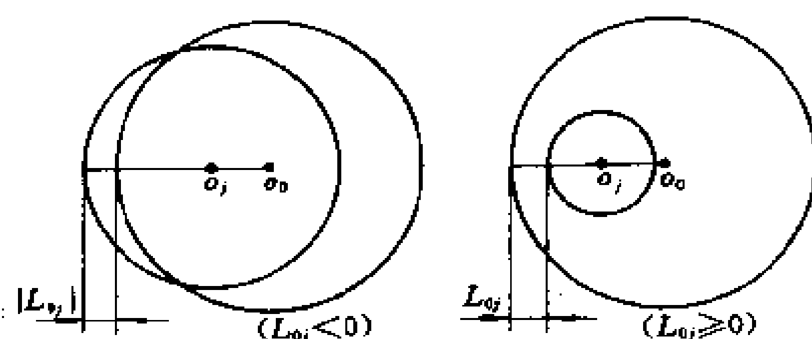


图 1-6　物体与容器的交的 Lebesque 测度

我们将容器外部看作带有一个空腔的弹性实体，并将此实体看作第 0 个物体，对于第 0 个物体与第 j 个 $(j = 1, 2, \cdots, N)$ 物体间的距离 L_{0j} 有类似于两物体间的距离 L_{ij} 的定义，如图 1-6 所示.

当两物体 i、j 的距离 $L_{ij} < 0$ 时，物体间产生挤压，因而产生弹性势能. 定义两物体间的弹性势能为

$$V_{ij}=\begin{cases}0, & 当\ L_{ij}\geqslant 0\ 时\\ L_{ij}^2, & 当\ L_{ij}<0\ 时\end{cases}$$

此式的弹性力学解释如下：当 $L_{ij}<0$ 时，$|L_{ij}|$ 表征了 i、j 两物体挤压变形的尺度，向弹性变形的势能正比于变形尺度的平方，而当 $L_{ij}\geqslant 0$ 时，i、j 两物体的交的测度为 0，没有发生挤压，因而弹性变形势能为 0.

由下式定义 N 个物体与空腔所构成系统的弹性势能 U：

$$U=\sum_{i=0}^{N-1}\sum_{j=i+1}^{N}U_{ij}$$

显然 U 是系统的状态的函数，即

$$U=U(x_1,y_1,z_1,\theta_1,\varphi_1,\psi_1,\cdots,x_N,y_N,z_N,\theta_N,\varphi_N,\psi_N)$$

而且 $U\geqslant 0$. 其中 $U>0$ 表明状态不满足 N 个物体在容器中放置的要求；$U=0$ 则表明满足放置的要求. 称使 $U=0$ 的状态为状态空间中的可行点.

利用下降算法可以寻找状态空间中的可行点. 一旦求出可行点，则状态即指明 N 个物体能放进容器中去的具体的姿态. 假若在计算相当长时间后仍算不出可行点，则或者是此问题无解，或者是由于以下三个原因造成此方法失灵：1°空间紧张；2°物体大小悬殊；3°物体个数多.

上述模型提供了一类 NP 难度问题的近似解法. 这里是对弹性力学进行模拟. 对统计物理进行模拟的模拟退火算法参见文献[27].

六、移植

在科学研究中，往往能够将一个或几个学科领域中的理论和行之有效的研究方法、研究手段移用到其他领域当中去，为解决其他学科领域中存在的疑难问题提供启发和帮助. 这是由于自然界各种运动形式之间的相互联系与相互统一，决定了各门自然科学之间的相互影响与相互渗透. 移植的特点是把问题的关键与已有

的规律和原理联系起来，与既存的事实联系起来，从而构成一个新的模型或深掘其本质的概念与思想．在类比和模拟中都包含有移植的成分，这里再看一个例子．

例 1-7 计算圆周率 π 的浦丰投针模型．

1977 年法国科学家浦丰（Buffon）利用几何概率研究了投针问题：在平面上画一些平行线，它们之间的距离都等于 a，向此平面任投一长度为 $l(l<a)$ 的针，用 x 表示针的中点到最近的一条平行线的距离，φ 表示针与平行线的交角，则针与平行线的位置关系如图 1-7 所示．

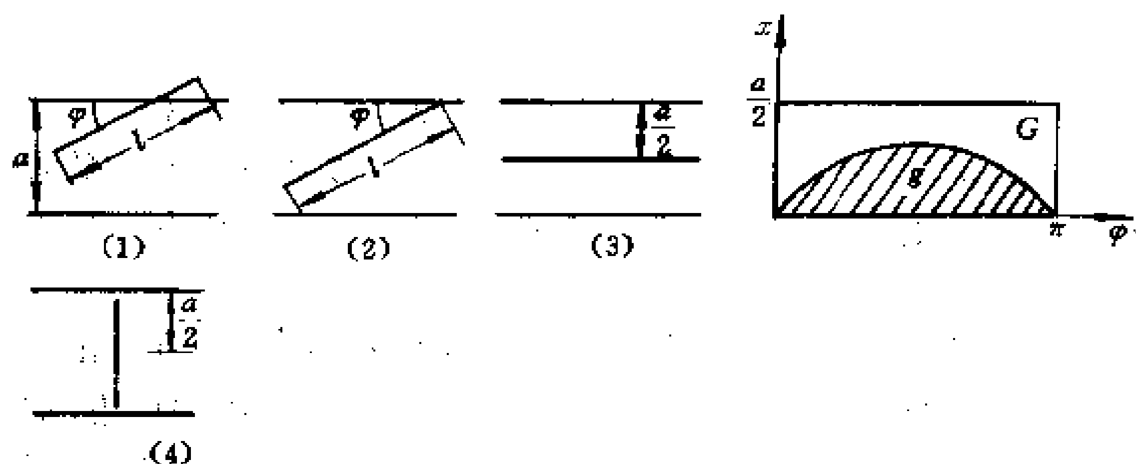


图 1-7 针与平行线的位置关系

显然 $0\leqslant x\leqslant \frac{a}{2}$，$0\leqslant\varphi\leqslant\pi$，以 G 表示边长为$\frac{a}{2}$及 π 的长方形，为使针与平行线相交，必须 $x\leqslant\frac{l}{2}\sin\varphi$，满足这个关系式的区域记为 g，则此针与任一平行线相交的概率为

$$p=\frac{g\text{ 的面积}}{G\text{ 的面积}}=\frac{\frac{1}{2}\int_0^{\pi} l\sin\varphi\,\mathrm{d}\varphi}{\frac{1}{2}a\pi}=\frac{2l}{\pi a}$$

所以，

$$\pi=\frac{2l}{a}\cdot\frac{1}{p}$$

因此为计算 π 的近似值，可在投针试验中记 N 为试验次数，

N_i 为针与平行线相交的次数，由大数定理：$\forall\ \epsilon>0$，$\lim\limits_{N\to\infty}P\{|\frac{N_i}{N}-p|\geq\epsilon\}=0$，可将$\frac{N_i}{N}$作为 p 的估计值，由此可得 π 的估计值.

此例采用的概率方法，随着电子计算机的发展，建立了一种新的方法——蒙特卡洛(Monte-Carlo)方法，并得到广泛的应用[16].

§1.3 观察力和想象力的培养

就数学成果的表述而论，数学以严密的演绎系统为其特征，然而数学发现的思维形式(包括产生、发展直到完善)却远非单纯的逻辑思维. 构造数学模型是一种创造性的工作，需要想象力、直觉和灵感(顿悟)这些非逻辑思维.

所谓想象，是人们对头脑中感知的形象(或称表象)进行加工创造新形象的心理活动，它不是表象的简单再现，而是表象的夸张、升华、理想化的改造. 例如"神"是由人的想象产生的，神具有人的模样，但又胜过人.

想象是形象的，具有概括性的. 想象时呈现于头脑中的是一幅整体的图景，是从整体上对事物进行思考的. 当然它在局部和细节上可能是模糊的，从而带来想象的自由性和灵活性.

科学的发现常常受益于想象的创造性功能. 爱因斯坦说："想象力比知识更重要，因为知识是有限的，而想象力概括着世界上的一切，推动着进步，并且是知识的源泉. 严格地说，想象力是科学研究中的实在因素."

微积分的发现是 17 世纪最伟大的数学成果，它是牛顿和莱布尼兹在许多数学家长期研究求切线斜率、求瞬时速度和研究曲边形面积计算方法的基础上，通过想象形成了粗糙而可贵的最初思想，这种发现是基于几何的直观和物理见解，并不是逻辑推理的结果.

直觉思维是人脑对客观世界及其关系的一种非常直接的识别或猜想的心理状态.它不是对事物先作各方面的详尽分析,按部就班地运用逻辑推理,达到对事物的认识,而是从整体上对待对象,越过思考的中间阶段,直接接触到结论的一种心智活动.笛卡尔认为通过直觉就能发现作为推理起点的、无可怀疑而清晰明白的概念.莱布尼兹认为,通过直觉可以认识自明的真理.两点之间,直线段距离最短,是出于直觉的认识.开普勒发现行星的公转周期 T 和它与太阳之间的距离 D 有关系 $T^2=D^3$,是因为它先有一种直觉的信念——行星运动是和谐的,T 和 D 之间必有某种和谐的关系,在这种直觉信念的促使下,才去不懈追求而发现的.

直觉是一种瞬间的判断,它以头脑中保持的信息为基础,凭借人们已有的大量知识和经验.它虽然不含详尽的推理,但它是依据事物整体的、最突出的特征来作出大致判断的,虽然表现出逻辑的中断,但它却是理性思维的"凝炼".直觉的结果是一种猜测,尽管其正确性必须经过严格的证明,但它却往往能提示解决问题的途径.

彭加勒认为,数学的发明与创造,无非是一种"组合"的"选择"而已.即从已有的数学事实(概念、判断、变换、结构、理论等等)出发,可形成无穷无尽的组合,而数学家的工作,就是要在这无穷的组合中,选择出有用的组合,扬弃无用的组合.他认为,摆在我们面前有无数条可供选择的道路,逻辑方法只能告诉我们走这条路或那条路不会遇到障碍,但它却不能向我们指明哪条道路可以达到目的地.人们只能从远处瞭望目标,而瞭望的本领就是直觉.

灵感又称顿悟.提出灵感,不由得使人想起阿基米德从洗澡盆里发现浮力,顿悟到测定皇冠含金量的方法,也使人想起笛卡尔解析几何的萌芽思想,产生于早晨枕上初醒时的佳谈.灵感是一种高度复杂的思维活动,是人们在文学创作或科学研究活动中,因思想高度集中而突然表现出来的一种心智活动.

我国清代文艺理论家王国维,曾借用诗句描绘做学问的三个

阶段："昨夜西风凋碧树，独上高楼，望断天涯路；衣带渐宽终不悔，为伊消得人憔悴；众里寻他千百度，蓦然回首，那人却在灯火阑珊处."他指出了，第一要占有资料，看清方向；第二要刻意追求；第三产生灵感，顿然觉悟. 在建立数学模型的过程中，也有和它相似的阶段.

华罗庚说："天才在于勤奋，聪明在于积累."灵感不会从天而降，而是在一定知识储备的基础上，对疑难问题久经沉思之后的几种信息之间的突然沟通.

非逻辑思维是一种发散性思维，它与收敛性的逻辑思维相辅相成. 在数学建模过程中，非逻辑思维由于松散，自由，联想的方面广，有充分的灵活性，富有创造性，能直接接触到问题的结论，提供问题结论的一种猜测，作为研究的起点，往往能获得突破性的创新. 逻辑思维的类比和归纳能验证、细化这种猜测，而演绎、抽象等方法则使问题的解决建立在更为坚实可信的基础之上.

在数学建模中，联想活动中视角的变化往往产生意想不到的效果. 这一点，在摄影和美术中已有很多体会，"横看成岭侧成峰"正是它的生动写照.

下面是培养想象力的几个例子.

例1-8 某人家住T市，在外地工作，平时总是乘坐下午5:30到达T市的火车回去，他的妻子准时在车站接他. 有一天此人提前半小时下班，乘火车5:00到达T市，然后步行回家. 路上，他遇到了开车来接他的妻子，因此比平时早10分钟到家. 问此人一共步行了多少时间?

粗看起来，似乎问题的条件不够而无法求解. 仔细分析，该问题涉及到两个人，由于问题的叙述和提问都是站在"某人"的角度，而"某人"到达T市的时刻在变化，因此顺着问题正面叙述的角度看，条件不足. 我们发现，在考虑问题时，实际上忽视了另一个人的存在，这就是"妻子". 而"妻子"每天从家到火车站的时刻表是不变的，这就找到了一个"参照系". 转而站在"妻子"的角度观察问题.

若“妻子”在这一天像往常一样继续开车到火车站(到火车站时刻为 5：30)然后回家，那么他们就会在平时一样的时刻到家. 这说明提前的 10 分钟是没有去车站省下的. 即相遇时，妻子距火车站还有 5 分钟的路程，因此相遇时刻为 5：25. 这说明此人已步行了 25 分钟 . 如图 1-8 所示. 图中虚线为节省的 10 分钟路程.

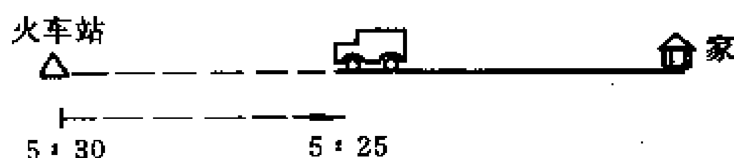


图 1-8

例 1-9 某人第一天上午 8：00 由 A 处出发，于下午 6：00 到达 B 处，第二天上午 8：00 他又由 B 处出发仍沿原路返回，并于下午 6：00 回到 A 处. 证明：途中至少存在一点，此人在两天中同一时刻到达该处.

将问题设想成甲乙两人在同一天分别从 A、B 两地相向而行，两人必在途中某处相遇，则结论的正确性就十分明显了.

习　　题

1. 什么是数学模型？它与我们在日常生活中接触到的照片、建筑模型以及美术中的写生、速写等绘画艺术有什么相似和不同？

2. 为了培养想象力、洞察力和判断力，考察对象时除了从正面分析外，还常常需要从侧面或反面思考. 试尽可能迅速地回答下列问题：

(1)37 支球队进行冠军争夺赛，每轮比赛中出场的每两支球队中的胜者及轮空者进入下一轮，直至比赛结束. 问共需进行多少场比赛？

(2)甲乙两站之间有电车相通，每隔 10 分钟甲乙两站相互发一趟车，但发车时刻不一定相同. 甲乙之间有一中间站丙站，某人每天在随机的时刻到达丙站，并搭乘最先经过丙站的那趟车，结果发现 100 天中约有 90 天到达甲站；仅约 10 天到达乙站. 问开往甲乙站的电车经过丙站的时刻表是如何安排的.

(3)一男孩和一女孩分别在离家 2000 米和 1000 米且方向相反的两所学校上学，每天同时放学后分别以 4 千米/小时和 2 千米/小时的速度步行回家. 一小狗以 6 千米/小时的速度由男孩处奔向女孩，又从女孩处奔向男孩，如此往返直至回到家中，问小狗奔波了多少路程？如果男孩和女孩上学时小狗也往返奔波在他们之间，问当他们到达学校时小狗在何处？

3. 人在雨中沿一直线从一处向另一处行进时，雨的速度已知，假设无风，问行人步行的速度多大才能使淋雨量最少？

第一部分　机理分析法

第二章　比例分析法

建立变量之间函数关系的方法很多，其中最基本最常用的方法之一是比例方法. 举例来说，如果以 $1:l$ 的比例作出某对象的比例模型，那么表面积的比例是 $1:l^2$，体积的比例是 $1:l^3$.

例 2-1　包装成本问题[6].

考虑像面粉、洗涤剂或果酱之类的产品，它们常常是包装后出售的. 注意到包装比较大的按每克计算的价格较低. 人们通常认为这是由于节省了包装和经营的成本的缘故. 或许有人会问，这是主要原因吗？是否还有其他重要因素？能否构造一个简单模型来分析？

我们研究的是产品成本如何随包装大小而变化的规律.

在产品销售过程中，有批发价和零售价等不同的价格，它反映了销售的不同阶段. 从研究批发价格入手，即零售商对该产品所偿付的价格. 计入批发价格的主要成本是：生产该产品的成本 a、包装该产品的成本 b、运输该产品的成本 c 和包装材料的成本 d.

产品成本显然随商业竞争和经营规模不同而变化，在这里研究的是销售过程中的粗略规律，因此忽略这些因素，集中考虑在原料和买卖过程的费用上. 设该产品成本 a 与所生产的货物量成正比，记为 $a \propto W$，其中 W 为产品重量.

包装成本取决于装包、封包以及装箱备运所需要的时间. 装包时间大致与体积(因而与重量)成比例，而对于体积在一定范围内的包装，后两部分时间相差不大. 于是对于某些正的常数 f 和

g ,$b \approx fW+g$.

运费可能同时取决于重量和体积，因为体积与装满的包的重量成比例，所以 $c \propto W$.

包装用材料的成本较为复杂，它取决于包装生产者必须偿付的各种成本，即必须考虑对于包装品生产者来说的 a、b、c 和 d. 设对那些盛装制造最后包装品的原料的容器，其成本忽略不计，则每件包装的成本取决于它的重量和体积．若所考虑包装的变动范围不太大，可认为各种体积的包装所用的包装材料相同．因此每件包装所消耗材料量（因而也是每件包装的重量）与所覆盖的表面积成正比．每件包装品的体积与包装品的表面积或体积成正比，它取决于摊平后运输（像纸板之类）还是成型后运输（像玻璃器皿之类）. 所以打包者的成本

$$d=hW+kS+m$$

其中，$h \geqslant 0$、$k>0$，$m>0$ 均为常数，S 是表面积.

现在将比例法中涉及的自变量化为一个自变量——重量. 假设各种包装品在几何形状上是大致相似的，体积几乎与线性尺度的立方成正比，表面积几乎与线性尺度的平方成正比，即 $V \propto l^3$，$S \propto l^2$. 所以 $S \propto V^{2/3}$. 由于 $V \propto W$，有 $S \propto W^{2/3}$. 于是每克的批发成本是

$$\frac{成本}{W}=\frac{a+b+c+d}{W}=n+pW^{-1/3}+\frac{q}{W}$$

其中，n、p、q 为正数．由此看出，当包装增大时，即每包内产品的重量 W 增大时，每克的成本下降.

进一步的分析可以看到，每克产品的成本下降速度

$$r=-\frac{\mathrm{d}(成本/W)}{\mathrm{d}W}=\frac{p}{3W^{4/3}}+\frac{q}{W^2}$$

这是 W 的减函数．因此当包装比较大时，每克的节省率增加得比较慢．总节省率为

$$rW=\frac{1}{3}pW^{-1/3}+qW^{-1}$$

也是 W 的减函数，其直观解释是：购买预先包装好的产品时，把小型包装的包装规格(体积)增大一倍，每克所节省的钱，倾向于比大型的包装规格增大一倍所节省的钱多．这里说"倾向于"是因为模型是粗糙的．然而在定性预测中往往很可靠．而验证上述解释也是很容易的，只须计算 $\left.\frac{\text{成本}}{W}\right|_{W_1}-\left.\frac{\text{成本}}{W}\right|_{W_2}$ 的值，其中 $W_2=2W_1$.

此模型可推广于零售价格，零售成本取决于批发价、销售成本和仓库成本，后两种成本具有 $HW+M$ 的形式，因此上述结论也适用于零售价格.

习　题

1. 研究零售价格情况下，单位重量的产品成本与包装的关系.

2. 设 $W_1<W_2<W_3$ 是一种包装产品的不同包装的重量，C_1、C_2 和 C_3 是包装产品的每克成本．试推导下面结果

$$\frac{C_1-C_2}{W_2-W_1}=\frac{q}{W_1W_2}+\frac{p}{W_1{}^{1/3}W_2+(W_1W_2)^{2/3}+W_1W_2{}^{1/3}}>\frac{C_2-C_3}{W_3-W_2}$$

为什么说这结果类似于"r 是 W 的减函数"呢？

3. 对于许多普通物体(例如行进中的汽车和自由落体等)，大气阻力大致与 Sv^2 成正例，其中 S 是表面积，v 是速度.

(a)若 v 是落体的末速度，证明：对于类似比例的物体，有 $v\propto m^{11/3}$.

(b)证明，当与地面相撞时必然转换为某种其他能量形式的单位面积动能与 m 成正比.

(c)讨论落在不同大小的动物身上时有什么影响(提示：较大的动物的骨胳较粗).

4. 某校有三个系联合成立学生会，(1)试确定学生会席位分配方案．(2)若甲系有 100 名学生，乙系 60 名，丙系 40 名，学生会设 20 个席位，分配方案如何？(3)若丙系有 3 名学生转入甲系，3 名学生转入乙系，分配方案有何变化？(4)若在第(3)问中将学生会席位增加一席呢？(5)试确定一数量指标衡量席位分配的公平性，并以此检查(1)～(4).

5. 人在匀速行走(速度固定)时步长多大最省劲？把人行走时作的功看

作是人体重心的势能和两脚运动的动能之和.试在此基础上建立数学模型并对结果进行评价.



第三章　代数方法

高速电子计算机的出现导致了数学的离散化趋势．所谓离散化，主要指两方面的工作，其一是，把连续模型离散化，以便作数值处理，其二是，把实际问题直接抽象成离散的数据、符号和图形，然后以离散数学为主要工具来解决．本章讨论的主要是后者．这里涉及到的一些问题，最终都归结为集合和线性代数问题，可以说代数方法是求解离散问题的主要方法．

§3.1　状态转移

状态转移问题分为确定性状态转移和随机性状态转移两种，通常将确定性的问题仍称为状态转移，而将随机性的称为马氏链问题．

一、确定性状态转移

所谓确定性状态转移问题，讨论的是在一定条件下，系统由一状态转移到另一状态是否可能，如果可能转移的话，应如何具体实现．

例 3-1　安全渡河问题．

这里是一个智力游戏．三名商人各带着一名随从，要乘一只小船过河．这只小船最多只能容纳两个人．随从们密约，在河的任一岸，一旦他们的人数比商人多，就杀人取货．但是，如何乘船的大权操纵在商人们手中．商人们已获知了这项密约．请你为商人们制定一个安全过河的方案．

这是一个状态转移问题．这里希望能得到一个规格化的方法，可以机械地得到结果，从而具有推广意义．

记渡河过程中此岸的商人数为 $x(x=0,1,2,3)$，随从数为 y $(y=0,1,2,3)$，于是此岸的状态可用向量(x,y)表示，显然共有 $4\times4=16$ 种可能的状态．但是，其中只有以下一些状态对商人是安全的：

$(3,y)$(其中 $y=0,1,2,3$)，表示商人全在此岸；

$(0,y)$(其中 $y=0,1,2,3$)，表示商人全在对岸；

(x,y)(其中 $x=y=1,2$)，表示两岸的商人和随从一样多．这些状态的集合称为允许状态集合，记作

$$S=\{(0,y)\mid y=0,1,2,3;\\ (3,y)\mid y=0,1,2,3;\\ (x,y)\mid x=y=1,2\}.$$

状态转移需经状态运算来实现．摆一次渡就可以改变现有状态．这种状态运算正是我们要选择的策略，也称为决策，用向量(x,y)表示，即 x 名商人和 y 名随从乘船．显然允许决策集合为

$$D=\{(x,y)\mid 1\leqslant x+y\leqslant 2\}.$$

小船从此岸到对岸或从对岸到此岸的每一次航行，都造成状态的一次转移．用 $s_1(x,y)$，$s_2(x,y)$，… 表示状态的变化过程，其中 $s_i\in S$．用 $d_i(x,y)$表示状态 $S_i(x,y)$下的状态，其中 $d_i\in D$．因为 i 为奇数时，决策 d_i 表示小船从此岸到对岸；i 为偶数时，d_i 表示小船从对岸到达此岸，所以状态转移满足下列关系：

$$s_{i+1}=s_i+(-1)^i d_i.$$

于是制定安全过河方案归结为这样一个数学问题：求决策 $d_i\in D$ $(i=1,2,\cdots)$，使状态 $s_i\in S$．按上述规则，由初始状态 $s_1(3,3)$，经 n 步转移到达 $s_n(0,0)$．当然 n 越小越好．

有了上面的模型，就可以编一段程序用计算机求解．如果在计算过程中出现循环，就说明问题无解．在没有计算机的条件下，当问题比较简单时，可作如下分析：

第一次渡河

$$s_1(3,3)-\begin{cases} d_1(1,0) \\ d_1(0,1) \\ d_1(1,1) \\ d_1(0,2) \\ d_1(2,0) \end{cases} = \begin{cases} s_2(2,3) & (\text{不可行}) \\ s_2(3,2) & (\text{可行}) \\ s_2(2,2) & (\text{可行}) \\ s_2(3,1) & (\text{可行}) \\ s_2(1,3) & (\text{不可行}) \end{cases}$$

第二次渡河

对 $d_1(0,1)$

$$s_2(3,2)+\begin{cases} d_2(1,0) \\ d_2(0,1) \\ d_2(1,1) \\ d_2(0,2) \\ d_2(2,0) \end{cases} = \begin{cases} \text{不可能} \\ \text{还原为原状态 } s_1(3,3) \\ \text{不可能} \\ \text{不可能} \\ \text{不可能} \end{cases}$$

由此可以看到，应排除单人渡河的方案.

对 $d_1(1,1)$

$$s_2(2,2)+\begin{cases} d_2(1,0) \\ d_2(0,1) \\ d_2(1,1) \\ d_2(0,2) \\ d_2(2,0) \end{cases} = \begin{cases} s_3(3,2) & \text{可行} \\ s_3(2,3) & \text{不可行} \\ s_3(3,3)=s_1(3,3) & \text{还原} \\ \text{不可能} & \\ \text{不可能} & \end{cases}$$

对 $d_1(0,2)$

$$s_2(3,1)+\begin{cases} d_1(1,0) \\ d_1(0,1) \\ d_1(1,1) \\ d_1(0,2) \\ d_1(2,0) \end{cases} = \begin{cases} \text{不可能} & \\ s_3(3,2) & \text{可行} \\ \text{不可能} & \\ \text{还原为原状态 } s_1(3,3) & \\ \text{不可能} & \end{cases}$$

因此，到第二次渡河时，可采取的策略为 $d_1(1,1)\to d_2(1,0)$，或 $d_1(0,2)\to d_2(0,1)$.

这一方法，实际上是一种搜索技术. 搜索技术是解决离散问

题．比如整数规划等问题的有力工具，特别是 DFS（Depth First Search）搜索法是一种用途很广的算法，在人工智能上，其基本思想经常用到．

DFS 不同于穷举法，其基本策略是“向前走，碰壁回头”，即一旦发现前面已是“此路不通”，立即回头，改换路径，而不是一条道走到底．这种“回溯”策略大大加快了搜索的速度．作为练习，请读者自己写出安全渡河问题的 *DFS* 算法的程序，并在计算机上实现．

类似进行以上分析，可寻找出渡河方案．

这一模型可用图解法方便地解出．在 xy 坐标系上画出图 3-1 表示的方格，允许状态集合 S 是有 x 点的 10 个方格顶点，允许决策是沿方格线移动一或二格，第奇数次规定向左或下方移动，第偶数次向右或向上方移动．图中分别用实线和虚线表示了从 $S_1(3,3)$开始的移动方案，最后到达 $s_{12}(0,0)$．

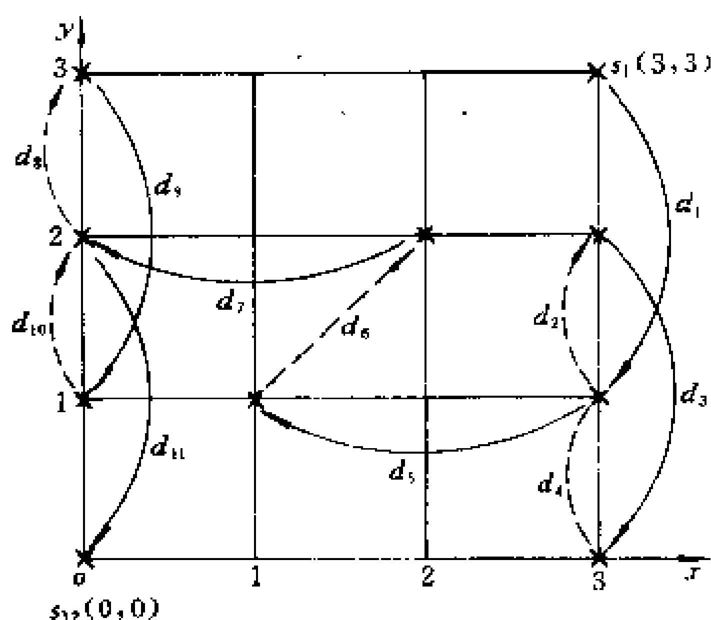


图 3-1　安全过河问题的图解法

二、随机状态转移——马氏链模型

马氏链是随机过程的一个特例，专门研究在无后效条件下时

间和状态均为离散的随机转移问题. 由于它在经济、生态、遗传、社会等领域中得到广泛的应用，并且很多确定性的状态转移问题也能用它处理，所以需要研究它的建模规律和处理方法. 当然，我们不打算深入讨论它的理论问题，对这一方面有兴趣的读者可以查阅有关书籍.

本节将先后讨论马氏链的随机型模型和确定型模型. 为了给没有学过马氏链的读者提供必要的基础，先通过两个例子简单介绍有关马氏链的基本知识.

例 3-2 某商店每月衡量一次经营情况，为简单起见，结果只用两种状态表示：$S=1$ 表示销路好，$S=2$ 表示销路坏. 假设下个月销路的好坏只与这个月的销路有关，与以前无关. 据经验，商店从本月处于状态 $S=i(i=1,2)$ 转到下月处于状态 $S=j(j=1,2)$ 的概率为 p_{ij}，用矩阵

$$p=(p_{ij})=\begin{bmatrix}0.5 & 0.5\\ 0.4 & 0.6\end{bmatrix}$$

表示. 如果开始时($t=0$)商店销路好($S=1$)，求 $t=1,2,\cdots$(t 以月为单位)商店处于各种状态的概率(如果开始时商店销路坏($S=2$)，请读者考虑.)

状态 $S=1,2$，以及转移概率 $p_{ij}(i,j=1,2)$ 可以直观地表示如图 3-2. 以 $a_i(t)(i=1,2;t=0,1,2,\cdots)$ 表示第 t 月商店处于状态 i 的概率. 因为 $a_i(t+1)$ 只与 $a_i(t)$ 有关，利用全概率公式有

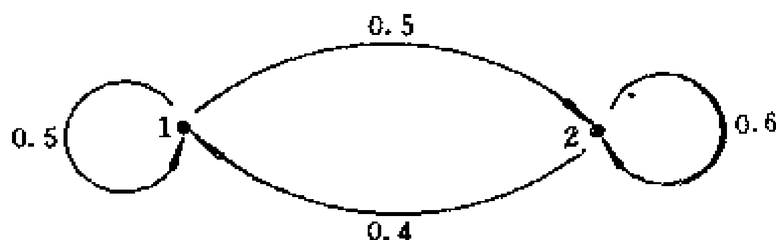


图 3-2　商店经营状态及转移概率示意图

$$\begin{cases}a_1(t+1)=a_1(t)p_{11}+a_2(t)p_{21}\\ a_2(t+1)=a_1(t)p_{12}+a_2(t)p_{22}\end{cases}\tag{1}$$

记行向量 $a(t)=(a_1(t),a_2(t))$，则(1)式可简记为

或
$$\begin{cases} a(t+1)=a(t)p \\ a(t)=a(0)p^t \end{cases} \tag{2}$$

若 $t=0$ 时商店销路好，则 $a_1(0)=1, a_2(0)=0$，由(2)式计算出的 $a(t)$ 如表 3-1 所示.

表 3-1 初值为(1,0)的 $a(t)$ 值

t	0	1	2	3	… →	∞
$a_1(t)$	1	0.5	0.45	0.445	… →	4/9
$a_2(t)$	0	0.5	0.55	0.555	… →	5/9

若 $t=0$ 时，销路坏，则 $a(0)=(0,1)$，由(2)计算出的 $a(t)$ 值如表 3-2 所示.

表 3-2 初值为(0,1)的 $a(t)$ 值

t	0	1	2	3	… →	∞
$a_1(t)$	0	0.4	0.44	0.444	… →	4/9
$a_2(t)$	1	0.6	0.56	0.556	… →	5/9

我们看到，不论商店处于哪种初始状态，$t\to\infty$ 时，商店处于两种状态的概率分别是 4/9 和 5/9，即时间充分长以后，商店的经营情况与初始状态无关.

由此，可以将马氏链的定义简述如下：一般地，如果按照过程的发展，时间可以离散化为 $t=0,1,2,\cdots$，对每个 t，描述过程的状态也可以离散化为 $1,2,\cdots,n$（有限个）. 从时刻 t 的状态 i 转移到 $t+1$ 时刻的状态 j 的概率是 p_{ij}. 若时刻 t 时，过程处于状态 $s_i(i=1,2,\cdots,n)$ 的概率是 $a_i(t)$，那么当 $a_i(t+1)=\sum_{j=1}^{n} a_j(t)p_{ij}(i=1,2,\cdots,n)$ 时，我们称这样的状态随机转移过程为马氏链. 显然，马氏链中时刻 $t+1$ 时过程所处各状态的概率只与时刻 t 时所处状态的

概率和转移概率有关，而与 $t-1$ 及 $t-1$ 以前时刻的状态无关. 所谓马氏性（马尔可夫性）或称无后效性，通俗地说，就是“已知现在，将来与过去无关”. 向量 $a(t)=(a_1(t),\cdots,a_n(t))$ 称为状态概率向量，$P=(p_{ij})$ 称为转移概率矩阵. $a_i(t)$ 和 p_{ij} 满足关系式

$$\sum_{i=1}^{n} a_i(t)=1 \qquad (t=0,1,2,\cdots)$$

$$p_{ij}\geqslant 0,\ \sum_{i=1}^{n} p_{ij}=1 \qquad (i=1,2,\cdots,n)$$

例 3-3 是马氏链的一个特例——正则链的实例. 所谓正则链，即如果对于任意的状态 i 和 j（i、j 可以相同），都存在正整数 k，使过程从状态 i 出发，经 k 步转移到状态 j，那么这个马氏链称为正则链.

由 $\boldsymbol{a}(t)=\boldsymbol{a}(0)p^t$ 知，上述条件等价于，存在正整数 k，使 $p^k>0$（指 p^k 的每一元素都大于零）.

可以证明，对于正则链存在一个极限状态概率向量 $\boldsymbol{W}=(W_1,W_2,\cdots,W_n)$，又称为稳定概率分布，使当 $t\to\infty$ 时，$\boldsymbol{a}(t)\to\boldsymbol{W}$，且 $\boldsymbol{W}$ 与初始状态概率向量 $\boldsymbol{a}(0)$ 无关.

对基本方程 $\boldsymbol{a}(t+1)=\boldsymbol{a}(t)$，两边令 $t\to\infty$，有

$$\boldsymbol{W}p=\boldsymbol{W}，其中 \sum_{i=1}^{n} W_i=1$$

可以解出 $\boldsymbol{W}$.

对于例 3-1，由

$$(W_1,W_2)\begin{pmatrix}0.5 & 0.5\\ 0.4 & 0.6\end{pmatrix}=(W_1,W_2)$$

$$W_1+W_2=1$$

有
$$\begin{cases}-0.5W_1+0.4W_2=0\\ W_1+W_2=1\end{cases}$$

解得 $\qquad W_1=4/9,\qquad W_2=5/9$

下面的例子，代表了另一种典型的马氏链.

例 3-4 讨论微量元素磷在如下过程中的转移情况. 磷在土壤中记作状态 1,在草、牛、羊等生物体中记为状态 2,在上述系统外记为状态 3. 转移概率 P_{ij} 可以理解为状态之中的磷转移到状态 j 中去的比例,如图 3-3 所示.

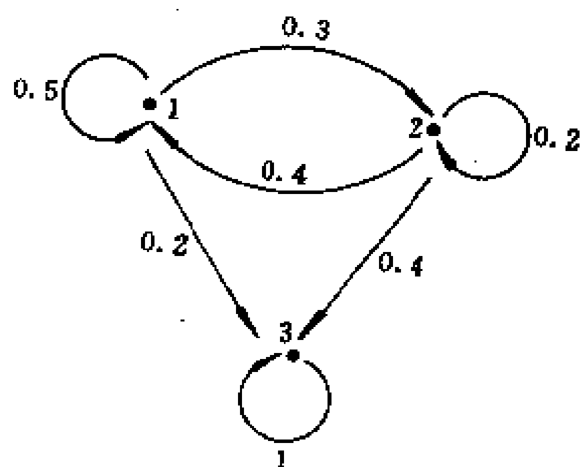


图 3-3 磷的状态转移

例如 $p_{21}=0.4$ 表示因草的枯死、牛羊排泄,磷又回到土壤中的比例为 0.4;$p_{33}=1$ 表示磷一旦移到系统(土壤、草、牛羊)之外,就不再进入系统,等等.

由图 3-3 知,转移概率矩阵为

$$p=\begin{bmatrix}0.5 & 0.3 & 0.2\\ 0.4 & 0.2 & 0.4\\ 0 & 0 & 1\end{bmatrix}$$

状态概率向量 $a(t)=(a_1(t),a_2(t),a_3(t))$ 可以理解为磷处于三种状态的比例. 状态转移过程具有无后效性,可用基本方程 $a(t+1)=a(t)p$ 计算.

若 $t=0$ 时磷全在土壤中,即 $a(0)=(1,0,0)$,则计算结果如表 3-3 所示.

我们看到 $t\to\infty$ 时,$a(t)\to(0,0,1)$,表示磷终将移出系统.

如果 $t=0$ 时,磷全部在系统外,即 $a(t)=(0,0,1)$,则对于任意的 t,总有 $a(t)=(0,0,1)$,表示磷一旦进入状态 3,就永远不会转移到其他状态.

在马氏链中，称 $p_{ii}=1$ 的状态 i 为吸收状态，例如例 3-2 中的状态 3. 如果马氏链至少含有一个吸收状态，并且从每一个非吸收状态出发都可以到达某个吸收状态，那么这个马氏链称为吸收链.

表 3-3 初值为(1,0,0)的 $a(t)$ 值

t	0	1	2	3	…	10	→	∞
$a_1(t)$	1	0.5	0.37	0.27	…	0.027	→	0
$a_2(t)$	0	0.3	0.21	0.15	…	0.015	→	0
$a_3(t)$	0	0.2	0.42	0.58	…	0.958	→	1

含有 m 个吸收状态和 $(n-m)$ 个非吸收状态的吸收链，其转移概率矩阵的标准形式为

$$p_{n\times n}=\begin{bmatrix} I_{m\times m} & 0 \\ R & Q_{(n-m)\times(n-m)} \end{bmatrix}$$

其中，矩阵 R 中含有非零元素，I 为单位阵.

例 3-2 中的转移概率矩阵的标准形式为

$$p=\begin{matrix} 3 \\ 1 \\ 2 \end{matrix}\begin{bmatrix} 1 & 0 & 0 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.4 & 0.2 \end{bmatrix}$$
$$\begin{matrix} 3 & \quad 1 & \quad 2 \end{matrix}$$

对于具有标准转移概率矩阵的吸收链，可以证明具有下列性质：

1° $t\to\infty$ 时，$Q^t\to 0$

2° $(I-Q)$ 可逆，且 $(I-Q)^{-1}=\sum_{t=0}^{\infty}Q^t$

3° 记 $(I-Q)^{-1}=N$，则从非吸收状态 i 出发，被某个吸收状态吸收之前的平均转移次数是 N 的第 i 行元素之和.

4° 记 $B=NR$，则从非吸收状态 i 出发，被吸收状态 j 吸收的概率是 B 的元素 b_{ij}.

在例 3-2 中， $Q=\begin{bmatrix} 0.5 & 0.3 \\ 0.4 & 0.2 \end{bmatrix}$,

$$I-Q=\begin{bmatrix} 0.5 & -0.3 \\ -0.4 & 0.8 \end{bmatrix}$$

所以

$$N=(I-Q)^{-1}=\frac{\begin{pmatrix} 0.8 & 0.3 \\ 0.4 & 0.5 \end{pmatrix}}{\begin{vmatrix} 0.5 & -0.3 \\ -0.4 & 0.8 \end{vmatrix}}=\begin{bmatrix} \frac{80}{28} & \frac{30}{28} \\ \frac{40}{28} & \frac{50}{28} \end{bmatrix}$$

因此,从非吸收状态 1 出发,到被吸收状态 3 吸收之前,平均转移次数为$\frac{80}{28}+\frac{30}{28}=3.93$;从非吸收状态 2 出发,到被吸收状态 3 吸收之前,平均转移次数为$\frac{40}{28}+\frac{50}{28}=3.21$.

$$B=NR=\begin{bmatrix} 2.86 & 1.07 \\ 1.43 & 1.79 \end{bmatrix}\begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix}=\begin{bmatrix} 1.00 \\ 1.00 \end{bmatrix}$$

因此,从非吸收状态 1 出发,被吸收状态 3 吸收的概率为 1;从非吸收状态 2 出发,被吸收状态 3 吸收的概率也是 1.

为进一步说明马氏链模型的应用,例 3-4 给出了遗传学研究中的一个例子.

例 3-5 基因遗传和近亲繁殖.

假设这样一种近亲繁殖情况:最初父母可以是优种、混种或劣种,它们有大量后代,从中随机选取一雄一雌进行交配,这样继续下去,分析后代的演变情况.

由于每次进行繁殖的是随机取来的一对,父和母都可能是 D、H 和 R 中的一种,组合起来有六种状态:DD、RR、DH、DR、HH 和 HR. 如 DH 状态表示取来一优种和一混种交配. 六种状态按上述顺序记作状态 1 到状态 6、状态转移概率 p_{ij}的含义为,交配状态 i 转移到后代状态 j 的概率,$i,j=1,2,\cdots,6$. 例如 P_{54}表示交配时的状态为 5(对应的状态为 HH,即取一混种和一混种交配),而后代为状态 4(对应状态为 DR),表示从父母为 HH 的后代中,任取一对为 DR 的概率.

下面以父母为HH状态为例，计算转移概率. 因为父母基因为dr和dr，所以从后代中选到优种D(基因为dd)的概率为$P(D)=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$，选到混种$H$(基因为$dr$)的概率为$P(H)=P(dr)+P(rd)=\frac{1}{2}\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{2}=\frac{1}{2}$，选到劣种$R$(基因为$rr$)的概率为$P(R)=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$. 故，当父母为$HH$时，从后代中任选一对为$DD$的概率$P(DD)=\frac{1}{4}\times\frac{1}{4}=\frac{1}{16}$，同理$P(RR)=\frac{1}{4}\times\frac{1}{4}=\frac{1}{16}$，$P(DH)=\frac{1}{4}\times\frac{1}{2}+\frac{1}{4}\times\frac{1}{2}=\frac{1}{4}$，$P(DR)=\frac{1}{4}\times\frac{1}{4}+\frac{1}{4}\times\frac{1}{4}=\frac{1}{8}$，$P(HH)=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$，$P(HR)=\frac{1}{2}\times\frac{1}{4}+\frac{1}{4}\times\frac{1}{2}=\frac{1}{4}$. 在转移概率矩阵中，相应为$P_{51},P_{52},\cdots,P_{56}$. 仿照此计算过程，得转移概率矩阵为

$$P=\begin{array}{c} \\ DD \\ RR \\ DH \\ DR \\ HH \\ HR \end{array}\begin{array}{c} \begin{array}{cccccc} DD & RR & DH & DR & HH & HR \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \frac{1}{16} & \frac{1}{16} & \frac{1}{4} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \end{array}$$

显然，DD和RR是两个吸收状态，这是一个吸收链. 它表明，在近亲繁殖下，不论最初选取的一对是什么，终将被DD和RR吸收，即变成全部是优种，或全是劣种. 而且一旦如此，就永远保持下去，不会变成其他种类.

利用转移概率矩阵P，计算可得

$$N=\begin{array}{c|cccc|} & DH & DR & HH & HR \\ DH & 8/3 & 1/6 & 4/3 & 2/3 \\ DR & 4/3 & 4/3 & 8/3 & 4/3 \\ HH & 4/3 & 1/3 & 8/3 & 4/3 \\ HR & 2/3 & 1/6 & 4/3 & 8/3 \end{array}\text{，行和为}\begin{bmatrix} 4\frac{5}{6} \\ 6\frac{2}{3} \\ 5\frac{2}{3} \\ 4\frac{5}{6} \end{bmatrix} \tag{3}$$

$$B=\begin{array}{c|cc|} & DD & RR \\ DH & 3/4 & 1/4 \\ DR & 1/2 & 1/2 \\ HH & 1/2 & 1/2 \\ HR & 1/4 & 3/4 \end{array} \tag{4}$$

根据 H 和 B 的性质，从(3)式可以看出，从状态 DH 出发，平均经过 $4\frac{5}{6}$代，后代就会全变成优种或劣种. 从(4)式可以看出，从状态 DH 出发，后代最终全变成优种的概率是 3/4，等等.

如果为了某种实际目的，不希望生物的后代全是优种或劣种，那么经过(3)式的结果说明，在近亲繁殖下，大约经过多少代应该重新选种.

例 3-6 联系网络中的成员地位度量.

在图 3-4 所示的警察联系网络中，在无线电通讯汽车中的警察 r，能把消息送到中心调度员 d，而中心调度员也能把消息传给他；巡警 c_1 和 c_2 可以传递消息给值班警察 s，值班警察却不能直接传递消息给巡警，只能要求中心调度员打电话给无线电通讯汽车，要求无线电通讯汽车中的警察把消息传给巡警. 现在要求制定一个度量，以测量给定的人在联系网络中的重要性，由此测量在有代表性的一天或一年中，某个人身上的担子有多重.

采用有向图作为联系网络的数学模型. 若 x 可以传递消息给 y，就得到一条从 x 到 y 的有向弧. 如图 3-4 所示.

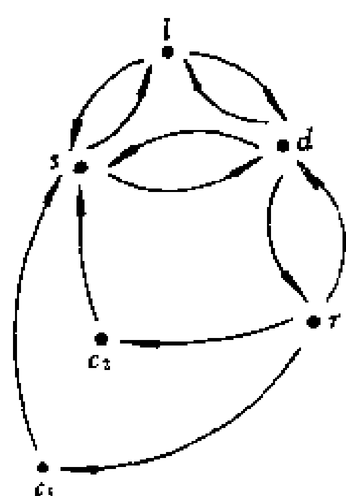


图 3-4　联系网络图示

其中 c_1、c_2 是两个巡警，d 是调度员，r 是无线电通讯汽车的驾驶员，s 是值班警察，l 是这一单位的负责人.

在这个图论模型中，假设消息是从一个地方起源——比如说图中的负责人，并假设接收者也送出一个他自己的消息. 从图中可以看出，任一成员发出的消息终将能传到其他任一个人. 由此可将图论模型扩展为一个马尔可夫链模型，而这是一个正则链. 另外，认为每个消息发出者将信息平均传递给他能直接传递的人，由此转移概率取为发出的信息量.

联系网络的转移矩阵如下：

$$P=\begin{array}{c} \\ l \\ s \\ d \\ c_1 \\ c_2 \\ r \end{array}\begin{array}{c} \begin{array}{cccccc} l & s & d & c_1 & c_2 & r \end{array} \\ \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix} \end{array}$$

设 $\boldsymbol{W}=(W_1,W_2,W_3,W_4,W_5,W_6)$ 表示该正则链的稳定概率分布向量，则由基本方程

$$\boldsymbol{W}=\boldsymbol{W}\boldsymbol{P},\quad \sum_{i=1}^{6}W_i=1$$

解得 $\boldsymbol{W}=\frac{1}{90}(22,26,27,3,3,9)$. 取该向量为联系网络中各个成员重要性的度量. 其理由在于：(1)这反映了在联系网络中，信息量

的稳定分布,从信息量的角度反映了网络中各成员的重要性;(2)这个度量与我们的直觉一致,例如我们的直觉是中心调度员最重要,而巡警最不重要,这在度量中得到反映;(3)交换联系网络中同等地位的成员,如 c_1 和 c_2,相应马氏链模型中概率转移矩阵不变,相应的重要性度量也不变.

这个例子说明了某些确定性问题有时也可以转化为随机性问题来处理,这种模型的转换,效果常常是意想不到的.

§3.2 动物种群的增长

生态平衡问题日益引起人们的重视.根据动物种群繁殖的机理,建立数学模型,是生态学的重要内容.在本节,我们希望从已知的初始种群的大小,预测经过一段时间以后的种群大小.还希望由已知种群的初始年龄分布,找出经过一段时间以后的年龄分布.

为简化讨论,只考虑两性种群中的雌性.如果计算两性别的成员,只要雌雄间的比例保持不变,而且各种年龄的死亡率对两个性别都相同,那么同样的论证也是适用的.莱斯卫斯(Lesvis 1942)和莱斯利(Leslie 1945,1948)对这一模型首先进行了研究.将雌性分成等区间长的年龄类,假设生殖率和死亡率在同一年龄间隔之内保持不变,对不同的年龄类才不相同,得到具有离散年龄等级的离散时间模型.显然这种模型是不精确的,但若时间间隔短,近似程度就好.

1. 年龄分段

设在此种群中,任一雌性最大可达年龄为 L 年,从而把种群雌性分为 n 个年龄类,每类跨越的年龄均为 L/n 年,如表 3-4 所示.

2. 观察时间离散化

我们在离散的时间 $t_0,t_1,\cdots,t_k,\cdots$ 观察,这相当于隔若干时间观察一次种群.并使两个相邻的观察时间间隔的长度和年龄分类

区间的长度一致，因此有

$$t_0=0, t_1=L/n, \cdots, t_k=k_L/n, \cdots.$$

表 3-4

年龄类	年龄区间
1	$(0,L/n]$
2	$(L/n,2L/n]$
…	…
n	$((n-1)L/n,L]$

若记在时刻 t_k 观察时第 i 类的数目为 $x_i^{(k)}, i=1,2,\cdots,n$，则由对时间的假定，在没有死亡者时，应有 $x_{i+1}^{(k+1)}=x_i^{(k)}$. 记 $x^{(k)}=(x_1^{(k)}, x_2^{(k)},\cdots,x_n^{(k)})^T$ 记在时刻 t_k 的年龄分布向量，则 $x^{(0)}$ 表示初始年龄分布，其具体数据可由统计资料得到.

3. 状态的转移

由于出生、死亡和成熟（进入下一阶段），观察的 n 类的数目将不断变化，希望能定量地描述这三个过程，从而推断出初始分布向量是如何随时间演变的.

记 a_i 为在第 i 类中一个雌性生雌性的平均数，$i=1,2,\cdots,n$，记 b_i 为第 i 类中的雌性能活着并变成第 $i+1$ 类的平均数 $i=1,2,\cdots,n-1$. 显然 $a_i\geqslant 0, i=1,2,\cdots,n, 0<b_i\leqslant 1, i=1,2,\cdots,n-1$. 其中 b_i 不能为零，否则没有一个雌性能活过第 i 类. 假设至少有一个 a_i 为正，否则没有出生过程. 此外 a_i, b_i 可由统计资料获得，如果环境变化不大，a_i, b_i 也是变化不大的.

对应于 $a_i>0$ 的年龄类称为能生育年龄类. 显然在时刻 t_k，第 1 类中雌性数目等于在 $t_{k-1}\sim t_k$ 间各类雌性生雌性数目之和，即

$$x_1^{(k)}=\sum_{i=1}^{n}a_i x_i^{(k-1)}$$

在时刻 t_k，第 $i+1$ 类中雌性的数目就是在时刻 t_{k-1} 第 i 类雌性中

活到时刻 t_k 的数目，即

$$x_{i+1}^{(k)}=b_i x_i^{(k-1)} \qquad i=1,2,\cdots,n-1.$$

利用矩阵记号可表示为：

$$\begin{pmatrix} x_1^{(k)} \\ x_2^{(k)} \\ \vdots \\ x_n^{(k)} \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & \cdots & a_{n-1} & a_n \\ b_1 & 0 & \cdots & \cdots & 0 \\ 0 & b_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & b_{n-1} & 0 \end{pmatrix} \begin{pmatrix} x_1^{(k-1)} \\ x_2^{(k-1)} \\ \vdots \\ x_n^{(k-1)} \end{pmatrix}$$

或简记为：　　$X^{(k)}=LX^{(k-1)} \qquad k=1,2,\cdots$

其中

$$L=\begin{pmatrix} a_1 & a_2 & \cdots & a_{n-1} & a_n \\ b_1 & 0 & & 0 & 0 \\ & b_2 & & & \vdots \\ & & \ddots & & \\ & 0 & & b_{n-1} & 0 \end{pmatrix}$$

称为 Leslie 矩阵（或射影矩阵）.

为估计种群增长过程的动态趋向，首先研究状态转移矩阵 Leslie 矩阵的特征值和特征向量.

令 $p(\lambda)$ 为 Leslie 矩阵的特征多项式，则

$$p(\lambda)=|\lambda I-L|=\begin{vmatrix} \lambda-a_1 & -a_2 & -a_3 & \cdots & -a_{n-1} & -a_n \\ -b_1 & \lambda & \ddots & & 0 & \\ & -b_2 & \ddots & \ddots & & \\ & 0 & & \ddots & \ddots & \\ & & & & -b_{n-1} & \lambda \end{vmatrix}$$

$$=\lambda^n-a_1\lambda^{n-1}-a_2b_1\lambda^{n-2}-a_3b_1b_2\lambda^{n-3}-\cdots-a_nb_1b_2\cdots b_{n-1}.$$

当 $\lambda\neq 0$ 时特征方程可变形为

$$a_1\lambda^{n-1}+a_2b_1\lambda^{n-2}+\cdots+a_nb_1b_2\cdots b_{n-1}=\lambda^n$$

用 λ^n 除两边，有

$$\frac{a_1}{\lambda}+\frac{a_2b_1}{\lambda^2}+\cdots+\frac{a_nb_1\cdots b_{n-1}}{\lambda^n}=1$$

定义函数 $q(\lambda)=\frac{a_1}{\lambda}+\frac{a_2b_1}{\lambda^2}+\cdots+\frac{a_nb_1\cdots b_{n-1}}{\lambda^n}$

则 $p(\lambda)=0$ 等价于 $q(\lambda)=1$(对于 $\lambda\neq0$).

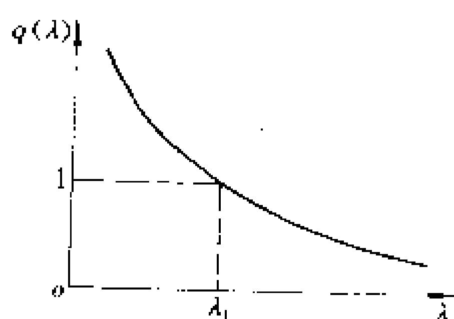


图 3-5

因为 $a_i\geqslant0(i=1,2,\cdots,n)$,$b_i>0(i=1,2,\cdots,n-1)$,故 $q(\lambda)$ 关于 $\lambda>0$ 单调下降,且 $\lambda\to0$,$q(\lambda)\to\infty$,而 $\lambda\to\infty$,$q(\lambda)\to0$,所以 $q(\lambda)$ 以 $\lambda=0$ 为垂直渐近线,$q(\lambda)$ 在第一象限的图形如图3-5 所示.
从而存在唯一的 λ,使 $q(\lambda)=1$,记此 λ 为 λ_1. 即矩阵 L 有唯一的正特征根且为单根.

下面求对应于 λ_1 的非零特征向量 $\boldsymbol{x}$. 由

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_{n-1} & a_n \\ b_1 & & & 0 & \\ & b_2 & & & \\ & & \ddots & & \\ 0 & & & b_{n-1} & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}=\lambda_1\begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}$$

有

$$\begin{cases} a_1x_1+a_2x_2+\cdots+a_nx_n=\lambda_1x_1 \\ x_2=\dfrac{b_1}{\lambda_1}x_1 \\ x_3=\dfrac{b_2}{\lambda_1}x_2 \\ \quad\cdots \\ x_n=\dfrac{b_{n-1}}{\lambda_1}x_{n-1} \end{cases}$$

令 $x_1=1$,则

$$\begin{cases} x_1 = 1 \\ x_2 = \frac{b_1}{\lambda_1} \\ x_3 = \frac{b_1 b_2}{\lambda_1^2} \\ \cdots \\ x_n = \frac{b_1 b_2 \cdots b_{n-1}}{\lambda_1^{n-1}} \end{cases}$$

显然此特征向量的所有元素为正，且它对应的特征子空间为一维，于是任何一个对应于 λ_1 的特征向量都是 x 的某一倍数.

综上所述，可知

定理 1 Leslie 矩阵 L 有唯一的正特征根 λ_1，它是单根，且相应的特征向量中有一个向量 x，其元素均为正数.

进一步讨论 λ_1 的性质，有

定理 2 设 λ_1 是 Leslie 矩阵 L 的唯一正特征根，则对 L 的任意其他特征根 λ_i（不论实的还是复的）都有 $|\lambda_i| \leqslant \lambda_1$.

实际上，设有一特征根 $\lambda = re^{i\theta}$，其模 $|\lambda| = r > \lambda_1$，因为 λ 为特征根，所以有

$$\frac{a_1}{\lambda} + \frac{a_2 b_1}{\lambda^2} + \cdots + \frac{a_n b_1 \cdots b_{n-1}}{\lambda^n} = 1$$

两边取模

$$1 = \left| \frac{a_1}{\lambda} + \frac{a_2 b_1}{\lambda^2} + \cdots + \frac{a_n b_1 \cdots b_{n-1}}{\lambda^n} \right| \leqslant \frac{a_1}{|\lambda|} + \frac{a_2 b_1}{|\lambda^2|} + \cdots + \frac{a_n b_1 \cdots b_{n-1}}{|\lambda^n|}$$
$$< \frac{a_1}{\lambda_1} + \cdots + \frac{a_n b_1 \cdots b_{n-1}}{\lambda_1^n} = 1$$

矛盾.

以上定理中称 λ_1 为 L 的占优特征根. 在下面的例子中我们看到若 λ_1 不是严格占优（即要求 $|\lambda_i| < \lambda_1$），生态系统可能出现周期性波动.

例 3-7 设

$$L=\begin{bmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix}$$

其特征多项式是

$$p(\lambda)=|\lambda I-L|=\begin{vmatrix} \lambda & 0 & -6 \\ -\frac{1}{2} & \lambda & 0 \\ 0 & -\frac{1}{3} & \lambda \end{vmatrix}=\lambda^3-1$$

三个特征根为 $\lambda_1=1,\lambda_2=-\frac{1}{2}+\frac{\sqrt{3}}{2}i,\lambda_3=-\frac{1}{2}-\frac{\sqrt{3}}{2}i$. 这三个特征根的模都是 1,从而这个唯一的正特征根 $\lambda_1=1$ 不是严格占优的. 但在这种情况下,矩阵有性质 $L^3=I$,这意味着对任意选定的初始年龄分布 $X^{(0)}$,有

$$X^{(0)}=X^{(3)}=X^{(6)}=\cdots=X^{(3k)}=\cdots \qquad (k=1,2,\cdots)$$

于是年龄分布向量以周期 3 发生振动(人口统计学家称之为人口波动),当 λ_1 为严格占优时,这种波动不会发生.

这里不讨论 λ_1 为严格占优的充分必要条件(详见文献[22]),而不加证明地给出一个充分条件.

定理 3 如果 Leslie 矩阵的第一行中有两个相邻的元素 a_i 和 a_{i+1} 不为零,则 L 的正特征根是严格占优的.

于是,如果种群有两个相邻的有生育能力的年龄类,则它的 Leslie 矩阵有一个严格占优的特征根. 实际上,只要年龄类的区间分得足够小,总会满足这个条件. 以后总假设定理三的条件满足.

现在来研究种群年龄分布的长期性态. 为使讨论简单,设 L 可对角化,且有 n 个特征根 $\lambda_1,\lambda_2,\cdots,\lambda_n$,以及对应于它们的线性无关的特征向量 $x_1,x_2,\cdots,x_n$,这些特征向量组成矩阵

$$P=[x_1,x_2,\cdots,x_n]$$

则 L 的对角化可由下式给出

$$L=P\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}P^{-1}$$

由此推得

$$L^k=P\begin{bmatrix}\lambda_1^k & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \lambda_n^k\end{bmatrix}P^{-1}$$

对于任何一个给定的初始年龄分布向量 $\boldsymbol{X}^{(0)}$，有

$$\boldsymbol{X}^{(k)}=L^k\boldsymbol{X}^{(0)}=P\begin{bmatrix}\lambda_1^k & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \lambda_n^k\end{bmatrix}P^{-1}\boldsymbol{X}^{(0)}$$

由于 λ_1 为严格占优的特征根，故 $\left|\dfrac{\lambda_i}{\lambda_1}\right|<1, i=2,\cdots,n$. 从而

$$\lim_{k\to\infty}\left(\frac{\lambda_i}{\lambda_1}\right)^k=0 \qquad i=2,3,\cdots,n.$$

由此知

$$\lim_{k\to\infty}\left\{\lambda_1\frac{1}{k}\boldsymbol{X}^{(k)}\right\}=P\begin{pmatrix}1 & 0 & \cdots & 0\\ 0 & & & \\ \vdots & & 0 & \\ 0 & & & \end{pmatrix}P^{-1}\boldsymbol{X}^{(0)}$$

记列向量 $P^{-1}\boldsymbol{X}^{(0)}$ 的第一个元素为 C，即

$$P^{-1}\boldsymbol{X}^{(0)}=(C,*,\cdots,*)^T.$$

则

$$P\begin{pmatrix}1 & 0 & \cdots & 0\\ 0 & & & \\ \vdots & & 0 & \\ 0 & & & \end{pmatrix}P^{-1}\boldsymbol{X}^{(0)}=[x_1,x_2,\cdots,x_n]\begin{pmatrix}c\\ 0\\ \vdots\\ 0\end{pmatrix}=c\boldsymbol{X}_1$$

其中,c 为正常数,仅与初始年龄分布有关.则

$$\lim_{k\to\infty}\{\lambda_1^{-k}X^{(k)}\}=cX_1$$

因此当 k 很大时,$X^{(k)}\approx c\lambda_1^kX_1$,而 $X^{(k-1)}\approx c\lambda_1^{k-1}X_1$,所以对充分大的 k,有 $X^{(k)}\approx\lambda_1X^{(k-1)}$. 这意味着对于充分大的时间,每一个年龄分布向量就是它前一期年龄分布向量的 λ_1 倍. 结果,在每一个年龄类中雌性的比例为常数,这个极限比例可由特征向量 X 定出.

例 3-8 设某类动物总数中雌性最大年龄为 15 岁,现把种群分为三类,每 5 年一类,设对此动物的 Leslie 矩阵为

$$L=\begin{pmatrix}0 & 4 & 3\\ \frac{1}{2} & 0 & 0\\ 0 & \frac{1}{4} & 0\end{pmatrix}$$

它的特征多项式 $P(\lambda)=|\lambda I-L|=\lambda^3-2\lambda-\frac{3}{8}$,其正特征根为 $\lambda_1=3/2$,所以 λ_1 相应的特征向量 X 为

$$X=\left(1,\frac{b_1}{\lambda_1},\frac{b_1b_2}{\lambda_1^2}\right)^T=\left(1,\frac{1}{3},\frac{1}{18}\right)^T$$

对于较大的 k,$X^{(k)}\approx\frac{3}{2}\cdot X^{(k-1)}$,从而知道在每一类中雌性的数目每 5 年增加 50%,并且由 $X^{(k)}=C\cdot\left(\frac{3}{2}\right)^kX=C\left(\frac{3}{2}\right)^k\left(1,\frac{1}{3},\frac{1}{18}\right)^T$ 知,这三类年龄中,雌性比例为

$$1\Big/\left(1+\frac{1}{3}+\frac{1}{18}\right)=18/25=0.72$$

$$\frac{1}{3}\Big/\left(1+\frac{1}{3}+\frac{1}{18}\right)=6/25=0.24$$

$$\frac{1}{18}\Big/\left(1+\frac{1}{3}+\frac{1}{18}\right)=1/25=0.04$$

下面对极限性态作进一步讨论. 对时间充分大时种群的年龄分布为

$$X^{(k)}=c\lambda_1^k X$$

这里有三种可能情况. (1)若 $\lambda_1>1$,则种群最终为增加;(2)若 $\lambda_1<1$,则种群数量最终为减少;(3)若 $\lambda_1=1$,则种群为稳定.

由此看到 $\lambda_1=1$ 在生态平衡中有特别的意义,因为它决定了种群的零增长. 因为 Leslie 矩阵的特征方程 $P(\lambda)=0$ 可写成 $q(\lambda)=1$,而

$$q(\lambda)=\frac{a_1}{\lambda_1}+\frac{a_2b_1}{\lambda_1^2}+\cdots+\frac{a_nb_1\cdots b_{n-1}}{\lambda_1^n}.$$

显然若令 $R=a_1+a_2b_1+\cdots+a_nb_1\cdots b_{n-1}$,则当 $\lambda_1=1$ 时,有 $q(1)=R=1$(因为 λ_1 满足方程 $q(\lambda)=1$). 反之,若 $R=1$,则等价于 $q(1)=1$,而 λ_1 是 $q(\lambda)=1$ 的唯一正实根,所以 $\lambda_1=1$ 的充要条件是 $R=1$. 称 R 是种群的纯生殖率. 因此,种群具有零增长当且仅当纯生殖率等于 1.

进一步可证: $R<1\Longleftrightarrow\lambda_1<1$, $R>1\Longleftrightarrow\lambda_1>1$. 实际上 $R<1\Longleftrightarrow q(1)<1\Longleftrightarrow q(1)<q(\lambda_1)\Longleftrightarrow\lambda_1<1$. 同理 $R>1\Longleftrightarrow q(1)>1\Longleftrightarrow q(1)>q(\lambda_1)\Longleftrightarrow\lambda_1>1$, 其中利用了 $q(\lambda)$ 在 $\lambda>0$ 内为单调减函数的性质.

这样不用计算 λ_1,而直接计算 R 即可得到种群数量增减的结论.

Leslie 的离散时间方法已用来模拟各种不同的生态种群,并发现此模型对预测未来人口状况有不可缺少的帮助.

在上述模型中,假设个体的生殖和死亡的机会都只是其年龄的函数,而不受种群大小的影响,即认为种群个体数目不会升高到使种群密度相关性开始发挥作用,Leslie(1959)考虑到这一点而提出了一种修正矩阵,并给出这种种群增长的一个数值例子.

在 Leslie 模型中 a_i、b_i 是由统计资料决定的常数(仅与年龄有关),但它们应是随机变量,具有各自的分布,因而应是一个随机过程问题. Sykes (1969) 和 Prollard (1966) 研究了它的随机变型. Prollard 指出此模型做出的预测并不明显地受离散时间间隔

长度变化的影响．他对比了对澳大利亚妇女每年增加率作出的两种估计，无论是按一年分组还是按五年分组，其估计都很接近．

Lefkovitch（1965）推广了 Leslie 模型，考虑用不相等的时间间隔，并在研究烟草甲虫的实验种群的增长中，证明了模型是有效的．

Skellam（1967）研究了在模拟种群周期性的季节改变而不是生活在恒温的不变条件下的 Leslie 矩阵，Williamson（1959）研究了两性比例不一致时的模型．Greville 和 Kegfitz（1974）研究了利用 Leslie 矩阵对过去情况的分析，这些工作都说明了模型的成功，也反映了随着建模要求的变化，模型假设的变化，模型也在不断发展．

§3.3　层次分析法

层次分析法（The Analytic Hierarchy Process，简称 AHP），是一种无结构的多准则决策方法，它将定性分析和定量分析相结合，把人们的思维过程层次化和数量化，在目标（因素）结构复杂且缺乏必要的数据情况下尤为实用．此方法自 70 年代美国运筹学家 Satry T. L. 提出以来，在实际中应用发展很快．

一、AHP 法原理

人们在日常生活中常常要做各种各样的决策．决策活动是人们进行选择或判断的一种思维活动．有的决策比较简单，而很多决策面临的常常是一个由相互关联、相互制约的众多因素构成的复杂系统，很难完全用定量的数学模型解决．人们在对决策问题的研究中，逐步认识到数学工具并非万能，决策中总会有大量因素无法定量地表示出来，而这正是软科学与通常的自然科学的区别．运筹学家们重新回到人的选择和判断上，认真研究决策思维的规律，AHP 正是在这种背景下提出的．

根据人的思维规律，面对复杂的选择问题，人们往往是将问题分解成各个组成因素，又将这些因素按支配关系分组形成递阶层次结构，通过两两比较的方式确定层次中诸因素的相对重要性，然后综合决策者的判断，确定决策方案相对重要性的总的排序，从而做出选择和判断．这一思维过程的关键是层次的划分，权重的确定和排序的并合规则．

例如，某工厂在扩大企业自主权后，厂领导考虑合理地使用企业留成的利润．可供选择的方案有：(1)发奖金；(2)扩建食堂、托儿所；(3)开办职工技校；(4)建图书馆；(5)引进新技术．领导在决策时，需要考虑到调动职工劳动生产积极性，提高职工文化水平和改善职工物质文化生活状况等方面．要作决策，必须对这些方案的优劣性进行排序，或者说必须决定每个方案在多目标下的权重．要处理这类复杂的决策问题，首先对问题所涉及的因素分类，构造一个各因素之间相互联结的层次结构模型．在上述问题中，因素分为三类：第一是目标类，即合理地使用今年企业留利××万元；第二是准则类，这是衡量目标能否实现的标准，如调动职工劳动积极性、提高企业的生产技术水平等等；第三是措施类，指实现目标的方案、方法、手段等等．按目标到措施自上而下地将各类因素之间的直接影响关系排列于不同层次，构成层次结构图．如图 3-6 所示．

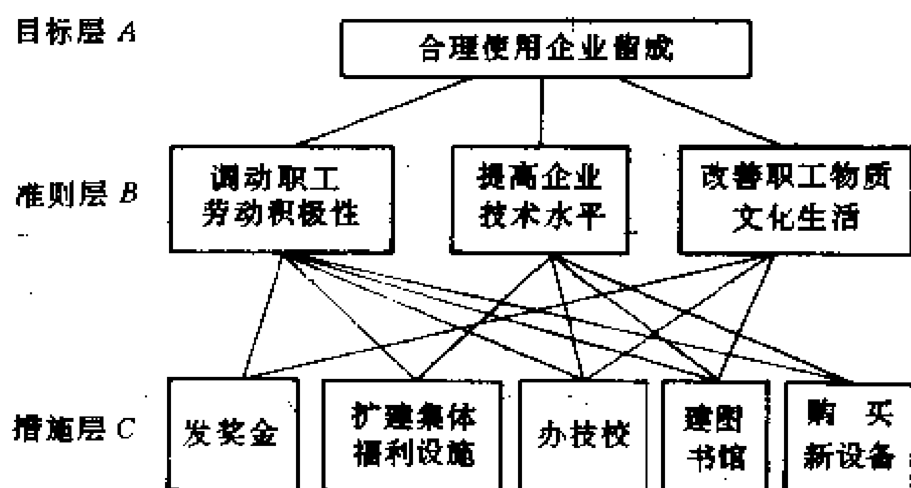


图 3-6

下面的工作就是通过成对比较的方法确定每一层的各因素的相对重要性的权重，直至计算出措施层各方案的相对权重，从而给出各方案的优劣次序.

决定各因素的相对权重的原理是这样的.

设有 n 件物体 $A_1, A_2, \cdots, A_n$，它们的重量分别为 $w_1, w_2, \cdots, w_n$，若将它们两两地比较重量，其比值可构成 $n \times n$ 矩阵

$$A=\begin{bmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \cdots & \cdots & \cdots & \cdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{bmatrix}$$

若用重量向量 $\boldsymbol{W}=(w_1, w_2, \cdots, w_n)^T$ 右乘 A 矩阵，则有

$$A\boldsymbol{W}=\begin{bmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ \cdots & \cdots & \cdots & \cdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{bmatrix}\begin{bmatrix} w_1 \\ \cdots \\ w_n \end{bmatrix}=n\begin{bmatrix} w_1 \\ \cdots \\ w_n \end{bmatrix}=n\boldsymbol{W}$$

即

$$(A-nI)\boldsymbol{W}=\boldsymbol{0}$$

可知 $\boldsymbol{W}$ 为特征向量，n 为特征值.

显然对于矩阵 A（称为成对比较阵或判断矩阵），有(1)$a_{ii}=1$；(2)$a_{ij}=1/a_{ji}(i,j=1,2,\cdots,n)$（互反性）；(3)$a_{ij}=a_{ik}/a_{kj}(i,j,k=1,2,\cdots,n)$（一致性）. 根据正矩阵的理论可证，$A$ 具有唯一非零的最大特征值 $\lambda_{max}=n$.

但是人们对复杂事物采用两两比较的方法获取成对比较阵时，不可能做到判断的完全一致性，而存在估计误差，这必然导致特征向量及特征值也有偏差. 用 $\overline{A}$ 和 $\overline{W}$ 表示相应的成对比较阵和特征向量，则问题由 $A\boldsymbol{W}=n\boldsymbol{W}$ 变成为 $\overline{A}\overline{\boldsymbol{W}}=\lambda_{max}\overline{\boldsymbol{W}}$. 这里 λ_{max} 是 $\overline{A}$ 的最大特征值，$\overline{W}$ 就是带有偏差的相对权重向量. 偏差是由判断的不相容引起的.

为了避免误差太大，这就要求衡量 $\overline{A}$ 矩阵的一致性. 注意到

当 A 矩阵完全一致时，由于 $a_{ii}=1$，$\sum_{i=1}^{n}\lambda_i=tr(A)=\sum_{i=1}^{n}a_{ii}=n$，存在唯一的非零 $\lambda=\lambda_{max}=n$，而当 $\bar{A}$ 存在判别不一致时，一般有 $\lambda_{max}\geqslant n$，这时

$$\lambda_{max}+\sum_{i\neq max}\lambda_i=\sum_{i=1}^{n}a_{ii}=n$$

故

$$\lambda_{max}-n=-\sum_{i\neq max}\lambda_i$$

取其平均值作为检验成对比较阵的一致性指标

$$CI=\frac{\lambda_{max}-n}{n-1}$$

当 $\lambda_{max}=n$ 时，$CI=0$，$\bar{A}$ 为完全一致；CI 值越大，成对比较阵的完全一致性越差.

由于 CI 中含有矩阵 $\bar{A}$ 的维数 n，一般 $\bar{A}$ 的维数 n 越大，判断的一致性将越差，应放宽对高维成对比较阵一致性的要求. Satty 提出采用随机性指标 CR，即对于固定的 n，随机地构造成对比较阵，其中 $a_{ii}=1$，$a_{ij}(i<j)$ 随机地从 $\frac{1}{9},\cdots,\frac{1}{2},1,\cdots,9$ 当中取一数. 可以认为这样的 $\bar{A}$ 是最不一致的. 用充分大的子样得到 $\bar{A}$ 的最大特征根的平均值 λ'_m，定义平均随机一致性指标为

$$RI=\frac{\lambda'_m-n}{n-1}$$

RI 值如表 3-5 所示

表 3-5

n	1	2	3	4	5	6	7	8	9	10
CR	0.00	0.00	0.58	0.96	1.12	1.24	1.32	1.41	1.45	1.49

当 $CR=\frac{CI}{CR}\leqslant 0.1$ 时，认为 $\bar{A}$ 的不一致性可以接受.

在构造 $\bar{A}$ 时，在成对比较中引入 1-9 的标度. 根据心理学家的研究，人们区分信息等级的极限能力为 7 ± 2. 标度 1，3，5，7，9

对应于 i 因素与 j 因素相比为同等重要，略为重要、比较重要、非常重要和绝对重要. 而 2,4,6,8 表示两判断之间的中间状态对应的标度值. 若 $a_{ij}=k(k=1,2,\cdots,9)$，则相应有 $a_{ji}=1/k(i\neq j)$. 当 $i=j$ 时，$a_{ij}=1$.

二、计算方法

在利用标度求得成对比较阵 $\overline{A}$ 后，AHP 方法的计算一般分为三步：1°计算特征向量 $\overline{W}$；2°计算最大特征值 λ_{max}；3°一致性检验. 由于 $\overline{A}$ 的近似性，$\overline{W}$ 和 λ_{max}的计算采用近似方法即可.

1. 特征向量 $\overline{W}$ 计算

介绍两种简化方法.

第一种方法，将 $\overline{A}$ 的各个列向量平均后，再标准化. 例如，若

$$\overline{A}=\begin{bmatrix}1 & 2 & 6\\ 1/2 & 1 & 4\\ 1/6 & 1/4 & 1\end{bmatrix}\xrightarrow{\text{按行求和}}\begin{bmatrix}1 & +2 & +6\\ 1/2 & +1 & +4\\ 1/6 & +4 & +1\end{bmatrix}\xrightarrow{\text{平均}}\begin{bmatrix}3\\ 1.83\\ 0.47\end{bmatrix}$$

$$\xrightarrow{\text{标准化}}\begin{bmatrix}0.57\\ 0.35\\ 0.08\end{bmatrix}=\overline{W}.$$

第二种方法，将 $\overline{A}$ 的各个列向量先标准化，再平均. 例如

$$\overline{A}=\begin{bmatrix}1 & 2 & 6\\ 1/2 & 1 & 4\\ 1/6 & 1/4 & 1\end{bmatrix}\xrightarrow{\text{按列标准化}}\begin{bmatrix}0.60 & 0.62 & 0.55\\ 0.30 & 0.31 & 0.36\\ 0.10 & 0.07 & 0.09\end{bmatrix}$$

$$\xrightarrow{\text{按行求和}}\begin{bmatrix}1.77\\ 0.97\\ 0.26\end{bmatrix}\xrightarrow{\text{平均}}\begin{bmatrix}0.59\\ 0.32\\ 0.09\end{bmatrix}=\overline{W}.$$

2. 计算 $\overline{A}$ 的最大特征值

$$\lambda_{max}=\frac{1}{n}\sum_{i=1}^{n}\frac{(A\overline{W})_i}{\overline{W}_i}$$

其中，$(A\overline{W})_i$ 为 $A\overline{W}$ 的第 i 个元素. 以上述 $\overline{A}$ 为例，利用方法一的结果.

$$A\overline{W}=\begin{bmatrix}1 & 2 & 6\\ 1/2 & 1 & 4\\ 1/6 & 1/4 & 1\end{bmatrix}\begin{bmatrix}0.57\\ 0.35\\ 0.08\end{bmatrix}=\begin{bmatrix}1.75\\ 0.955\\ 0.26\end{bmatrix}$$

$$\lambda_{max}=\frac{1}{3}\left[\frac{1.75}{0.57}+\frac{0.955}{0.35}+\frac{0.26}{0.08}\right]=3.016.$$

用精确方法，如计算特征值的幂法可得到 $\overline{A}$ 的最大特征值 $\lambda_{max}=3.01$，特征向量 $W=(0.588,0.322,0.090)^T$.

3. 一致性检验

$n=3$，查表知 $RI=0.58$，此时

$$CI=\frac{3.016-3}{3-1}=0.008$$

$$CR=\frac{CI}{CR}=\frac{0.008}{0.58}=0.014<0.1$$

$\overline{A}$ 的不一致性可以接受.

4. 层次总排序及一致性检验

上述过程中求出的 $\overline{W}$ 是同一层次中相应元素对于上一层次中的某个因素相对重要性的排序权值，这称为层次单排序. 若模型由多层次构成，计算同一层次所有因素对于最高层（总目标）相对重要性的排序权值，称为层次总排序. 这一过程是由最高层到最低层逐层进行的. 设上一层次 A 包含 m 个因素 $A_1,A_2,\cdots,A_m$，其层次总排序权值分别为 $a_1,a_2,\cdots,a_m$，下一层次 B 包含 n 个因素 B_1，$B_2,\cdots,B_n$，它们对于 A_j 的层次单排序权值分别为 $b_{1j},b_{2j},\cdots,b_{nj}$（当 B_k 与 A_j 无联系时，$b_{kj}=0$），此时 B 层 i 元素在层次总排序中的权值为 $W_i=\sum_{j=1}^{m}a_jb_{ij}(i=1,2,\cdots,n)$.

层次总排序也要进行一致性检验. 检验是从最高层到低层进行的. 设 B 层中的某些因素对 A_j 单排序的一致性指标为 CI_j，平均随机一致性指标 RI_j，则 B 层总排序随机一致性比率为

$$CR=\sum_{j=1}^{m}a_jCI_j\Big/\sum_{j=1}^{m}a_jRI_j.$$

当 $CR \leqslant 0$ 时，认为层次总排序结果具有满意的一致性.

三、实例

例 3-9 企业留成问题.

在企业留成的决策问题中，若构造的判断矩阵如下：

判断矩阵 $A-B$

$$
\begin{array}{c} A \\ B_1 \\ B_2 \\ B_3 \end{array}
\begin{array}{c} \begin{array}{ccc} B_1 & B_2 & B_3 \end{array} \\ \begin{bmatrix} 1 & \frac{1}{5} & \frac{1}{3} \\ 5 & 1 & 3 \\ 3 & \frac{1}{3} & 1 \end{bmatrix} \end{array}
$$

判断矩阵 B_1-C

$$
\begin{array}{c|ccccc} B_1 & C_1 & C_2 & C_3 & C_4 & C_5 \\ \hline C_1 & 1 & 3 & 5 & 4 & 7 \\ C_2 & \frac{1}{3} & 1 & 3 & 2 & 5 \\ C_3 & \frac{1}{5} & \frac{1}{3} & 1 & \frac{1}{2} & 2 \\ C_4 & \frac{1}{4} & \frac{1}{2} & 2 & 1 & 3 \\ C_5 & \frac{1}{7} & \frac{1}{5} & \frac{1}{2} & \frac{1}{3} & 1 \end{array}
$$

判断矩阵 B_2-C

$$
\begin{array}{c|cccc} B_2 & C_2 & C_3 & C_4 & C_5 \\ \hline C_2 & 1 & \frac{1}{7} & \frac{1}{3} & \frac{1}{5} \\ C_3 & 7 & 1 & 5 & 3 \\ C_4 & 3 & \frac{1}{5} & 1 & \frac{1}{3} \\ C_5 & 5 & \frac{1}{3} & 3 & 1 \end{array}
$$

判断矩阵 B_3-C

$$\begin{array}{c|cccc} B_3 & C_1 & C_2 & C_3 & C_4 \\ \hline C_1 & 1 & 1 & 3 & 3 \\ C_2 & 1 & 1 & 3 & 3 \\ C_3 & \frac{1}{3} & \frac{1}{3} & 1 & 1 \\ C_4 & \frac{1}{3} & \frac{1}{3} & 1 & 1 \end{array}$$

相应矩阵的层次单排序计算和一致性检验结果如下：

判断矩阵 $A-B$

$$\overline{W}=\begin{bmatrix}0.105\\0.637\\0.258\end{bmatrix} \qquad \begin{aligned}&\lambda_{\max}=3.038, CI=0.019,\\&RI=0.58, CR=0.033\end{aligned}$$

判断矩阵 B_1-C

$$\overline{W}^{(1)}=\begin{bmatrix}0.491\\0.232\\0.092\\0.138\\0.046\end{bmatrix} \qquad \begin{aligned}&\lambda_{\max}^{(1)}=5.126, CI_1=0.032\\&RI_1=1.12,\ CR_1=0.028\end{aligned}$$

判断矩阵 B_2-C

$$\overline{W}^{(2)}=\begin{bmatrix}0.053\\0.564\\0.118\\0.263\end{bmatrix} \qquad \begin{aligned}&\lambda_{\max}^{(2)}=4.117, CI_2=0.039\\&RI_2=0.90, CR_2=0.043\end{aligned}$$

判断矩阵 B_3-C

$$\overline{W}^{(3)}=\begin{bmatrix}0.406\\0.406\\0.094\\0.094\end{bmatrix} \qquad \begin{aligned}&\lambda_{\max}^{(3)}=4,\ CI_3=0\\&CR_3=0\end{aligned}$$

各个方案相对于目标后的总排序计算如表 3-6.

表 3-6　各个方案相对于目标层的总排序

层次 C ＼ 层次 B	B_1	B_2	B_3	层次 C 总排序 W
	0.105	0.637	0.258	
C_1	0.491	0	0.406	0.157
C_2	0.232	0.055	0.406	0.164
C_3	0.092	0.564	0.094	0.393
C_4	0.138	0.118	0.094	0.113
C_5	0.046	0.263	0	0.172

层次总排序一致性检验如下

$$CI=\sum_{i=1}^{5}B_iCI_i=0.105\times0.032+0.637\times0.039+0.258\times0$$
$$=0.028$$
$$RI=\sum_{i=1}^{3}B_iRI_i=0.105\times1.12+0.637\times0.90+0.258\times0.90$$
$$=0.933$$
$$CR=\frac{CI}{RI}=\frac{0.028}{0.933}=0.025<0.1$$

因此，层次总排序符合一致性要求．所提出的五种方案的相对优先排序为：开办职工技术学校（$W_3=0.393$）；引进新技术设备，进行企业改造（$W_5=0.072$）；扩建职工宿舍、食堂等福利设施（$W_2=0.164$）；发放奖金（$W_1=0.157$）和扩建图书馆俱乐部（$C_4=0.113$）．由此厂领导和职代会可根据上述分析结果，决定各种方案的实施先后次序，或决定分配企业留成利润的比例．

§3.4　组合优化与 NP 问题

组合数学作为离散数学这个“家族”的主要成员，在计算机的刺激下得到迅猛发展，并在一些以前与数学没有多大关联的学科，如化学、生物学和社会学等领域中得到广泛的应用．

组合数学研究的对象就是一组事物安排成各种模式的问题. 有两类问题经常出现:

1° 安排的存在性. 要把一组事物进行安排,使之满足某些条件. 当能否这样安排不是那么明显时,就需要讨论存在问题. 如果一种安排不总是可能的,那么在怎样的条件下,才能使所希望的安排办得到.

2° 安排的计数和分类. 如果安排是可能的,可以用许多方式来完成,人们可能要求计算这些方式的个数或者把它们进行分类.

与 1°连同出现的第三个组合问题是:

3° 研究一个已知安排. 在人们已经构造出了(可能是困难的)满足一定条件的安排之后,就可以研究这种安排的性质和结构了. 这样的结构可能涉及分类问题 2°,并且也可能牵涉到潜在的应用.

一般来说,组合数学与离散结构和关系的分析有关. 对于验证一些发现,数学归纳法是组合数学的主要工具之一. 另外还有容斥原理、鸽笼原理、递归关系以及生成函数等. 解决组合问题需要特殊的方法,其中经验是极为重要的. 例如 n 个人,n 项任务的安排,它共有 $n!$ 种不同方案,情况各异. 若对每项方案进行穷举,以 $n=100$ 为例,$100!=9.3\times10^{157}$,如果采用每秒可完成 10^7 个方案的高速电子计算机计算,也需要 2.96×10^{143}年. 显然,这种算法虽然在理论上可行,实际却办不到. 类似于这种问题的一类离散对象的问题被层出不穷地提了出来,于是非数值计算一类问题的算法研究,便成为近代组合数学研究的重要内容,被称为组合算法. 有一类优化问题具有某些组合意义,如零件加工的排序问题、旅行商问题、最短路径问题等等,因此,也称这类优化问题为组合优化问题.

求解组合问题最直接的方法是穷举法,利用计算机速度快和存贮量大的优点,在问题的规模很小时,可求出问题的全部解. 现在有很多有效的组合算法, 如 *DFS* 搜索法、分支定界法、隐枚举法、动态规划等,都属于一类不完全穷举法,另外还有大量近似算

法和特殊算法. 这里仅仅就几个比较典型的组合优化问题讨论问题的建模和求解.

一、动态规划模型

在生产和科学实验中,有一类活动的过程,由于它的特殊性,可将过程分为若干个互相联系的阶段,在它的每一阶段都需要作出决策,从而使整个过程达到最好的活动效果. 因此,各个阶段决策的选取不是任意确定的,它依赖于当前面临的状态,又影响以后的发展. 当各个阶段决策确定后,就组成了一个决策序列,因而也就决定了整个过程的一条活动路线. 这种把一个问题看作是一个前后关联具有链状结构的多阶段过程称为多阶段决策过程,也称为序贯决策过程,这种问题就称为多阶段决策问题. 动态规划是解决多阶段决策过程最优化的一种数学方法,是运筹学的一个分支.

我们通过最短路线问题说明动态规划的基本方法.

例 3-10 最短路线问题.

如图 3-7 所示,给定一个线路网络,两点之间连线上的数字表示两点间的距离(或费用),试求一条由 A 到 G 的铺管线路,使总距离为最短(或总费用最小).

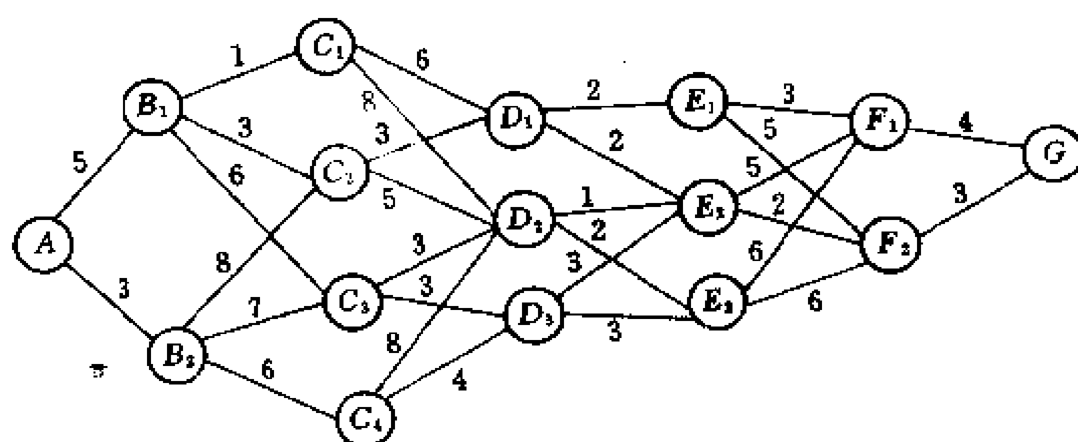


图 3-7

根据生活常识,最短路线有一个重要特征:如果由起点 A 经过 P 点而到达终点 G 的路线是一条最短路线,则由 P 点出发的这

段子路线，对于从 P 点出发到达终点的所有可能选择的不同路线来说，必定也是最短路线. 因为如果不是这样，则从点 P 到终点 G 有另一条距离更短的路线存在，把它和原来路线中由 A 到 P 的那部分连接起来，就会得到一条由 A 到 G 的新路线. 它比原来那条最短路线的距离要短些，这与假设矛盾，是不可能的.

根据最短路线这一特性，寻找最短路线的方法，就是从最后一段开始，用由后向前逐步递推的方法，求出各点到 G 点的最短路线，最后求得由 A 点到 G 点的最短路线. 在本题中，可分为 6 个阶段求解，取 K 为阶段变量，$k=1,2,\cdots,6$. 用 S_k 表示第 k 阶段的状态变量，例如 $S_3=\{c_1,c_2,c_3,c_4\}$；用 $u_k(S_k)$ 表示第 k 阶段处于 S_k 时的决策变量；用 $f_k(S_k)$ 表示从第 k 阶段的状态 S_k 开始到第 n 阶段的终止状态的过程，采取最优策略所得到的指标函数值；用 $d(S_k,S_{k-1})$ 表示状态 S_k 和 S_{k+1} 间的距离.

当 $k=6$ 时，由 F_1 到终点 G 只有一条路线，所以 $f_6(F_1)=4$. 同理，$f_6(F_2)=3$.

当 $k=5$ 时，出发点有 E_1,E_2,E_3 三个，若从 E_1 出发，有两种选择，一是到 F_1，另一个是到 F_2，则

$$f_5(E_1)=\min\begin{Bmatrix}d_5(E_1,F_1)+f_6(F_1)\\d_5(E_1,F_2)+f_6(F_2)\end{Bmatrix}=\min\begin{Bmatrix}3+4\\5+3\end{Bmatrix}=7$$

相应决策为 $u_5(E_1)=F_1$. 这说明，由 E_1 到终点 G 的最短路线为 7，其最短路线为 $E_1\to F_1\to G$.

同理，从 E_2 和 E_3 出发，有

$$f_5(E_2)=\min\begin{Bmatrix}d_5(E_2,F_1)+f_6(F_1)\\d_5(E_2,F_2)+f_6(F_2)\end{Bmatrix}=\min\begin{Bmatrix}5+4\\2+3\end{Bmatrix}=5$$

相应决策为 $u_5(E_2)=F_2$.

$$f_5(E_3)=\min\begin{Bmatrix}d_5(E_3,F_1)+f_6(F_1)\\d_5(E_3,F_2)+f_6(F_2)\end{Bmatrix}=\min\begin{Bmatrix}6+4\\6+3\end{Bmatrix}=9$$

相应决策为 $u_5(E_3)=F_2$.

类似可算得

当 $k=4$ 时，$f_4(D_1)=7$ $\qquad u_4(D_1)=E_2$

$f_4(D_2)=6$ $\qquad u_4(D_2)=E_2$

$f_4(D_3)=8$ $\qquad u_4(D_3)=E_2$

当 $k=3$ 时，$f_3(C_1)=13$ $\qquad u_3(C_1)=D_1$

$f_3(C_2)=10$ $\qquad u_3(C_2)=D_1$

$f_3(C_3)=9$ $\qquad u_3(C_3)=D_2$

$f_3(C_4)=12$ $\qquad u_3(C_4)=D_3$

当 $k=2$ 时，$f_2(B_1)=13$ $\qquad u_2(B_1)=C_2$

$f_2(B_2)=16$ $\qquad u_2(B_2)=C_3$

当 $k=1$ 时，出发点只有 A 一个点，

$$f_1(A)=\min\begin{Bmatrix} d_1(A,B_1)+f_2(B_1) \\ d_1(A,B_2)+f_2(B_2) \end{Bmatrix}=\min\begin{Bmatrix} 5+13 \\ 3+16 \end{Bmatrix}=18$$

相应决策 $u_1(A)=B_1$，于是得到从 A 到终点 G 的最短距离为 18.

为了找到最短路线，再按计算顺序反推，可求得最优决策函数序列 $\{u_k\}$，即由 $u_1(A)=B_1$，$u_2(B_1)=C_2$，$u_3(C_2)=D_1$，$u_4(D_1)=E_2$，$u_5(E_2)=F_2$，$u_6(F_2)=G$，组成一个最优策略，相应路线为

$$A\to B_1\to C_2\to D_1\to E_2\to F_2\to G$$

上述计算过程，利用了 k 阶段和 $k+1$ 阶段之间的递推关系：

$$\begin{cases} f_k(S_k)=\min\limits_{u_k\in D_k(S_k)}\{d_k(S_k,u_k(S_k))+f_{k+1}(u_k(S_k))\} \\ \qquad\qquad k=6,5,4,3,2,1 \\ f_7(S_7)=0 \quad (\text{或写成 } f_6(S_6)=d_6(S_6,G)) \end{cases}$$

其中 $D_k(S_k)$ 表示第 k 阶段从状态 S_k 出发的允许决策集合.

一般情况下，k 阶段与 $k+1$ 阶段的递推关系式可写作

$$f_k(S_k)=\mathop{\mathrm{opt}}\limits_{u_k\in D_k(S_k)}\{U_k(S_k,u_k(S_k))+f_{k+1}(u_k(S_k))\}$$

$$k=n,n-1,\cdots,2,1$$

边界条件为

$$f_{n+1}(S_{n+1})=0$$

这种递推关系式称为动态规划的基本方程. 从上例中我们看

到，将一个实际问题建立动态规划模型时，必须做到下面五点：

1° 将问题的过程划分成恰当的阶段.

2° 正确选择状态变量 S_k，使它既能描述过程的演变，又要满足无后效性.

3° 确定决策变量 u_k 及每阶段的允许决策集合 $D_k(S_k)$.

4° 正确写出状态转移方程.

5° 正确写出指标函数 $V_{k,n}$的关系，它应满足三个性质：第一，它是定义在全过程和所有 k 子过程上的数量函数；第二，它具有可分离性和递推关系，即

$$V_{k,n}(S_k,u_k,\cdots,S_{n-1})=\psi_k[S_k,u_k,V_{k-1,n}(S_{k-1},u_{k-1},\cdots,S_{n-1})];$$

第三，指标函数 $\psi_k(S_k,u_k,V_{k-1,n})$对于变量 $V_{k-1,n}$为严格单调，通常将指标函数取为过程所包含的各阶段的指标的和或积的形式.

下面是应用动态规划来解决组合优化问题的几个实例.

例 3-11 旅行商问题.

旅行商问题也称为货郎担问题，是运筹学中的一个著名问题：有一个串村走户的卖货郎，从某个村庄出发，通过若干个村庄一次且仅一次，最后又回到原来出发的村庄. 问应如何选择行走路线，总的行程最短.

这个问题的一般提法是：设有 n 个城市，以 $1,2,\cdots,n$ 表示，d_{ij} 表示从 i 城到 j 城的距离. 一个推销员从城市 1 出发到其他每个城市去一次且仅仅是一次，然后回到城市 1. 问他如何选择路线，使总的路程最短. 这个问题属于组合最优化问题，当 n 不太大时，利用动态规划方法求解是方便的.

由于规定推销员是从城市 1 开始的，设推销员走到 i 城，记

$N_i=\{2,3,\cdots,i-1,i+1,\cdots,n\}$表示由 1 城到 i 城的中间城市集合.

S 表示到达 i 城之前中途所经过的城市的集合，则有 $S<N_i$.

因此，可选取(i,S)作为描述过程的状态变量，决策为由一个城市走到另一个城市，并定义最优值函数 $f_k(i,S)$为从 1 城市开

始，经由 k 个中间城市的 S 集到 i 城的最短路线的距离，则可写出动态规划的递推关系为：

$$f_k(i,S)=\min_{j\in S}[f_{k-1}(j,S\setminus\{j\})+d_{ji}]$$

$$(k=1,2,\cdots,n-1,i=2,3,\cdots,n,S\subset N_i)$$

边界条件为 $f_0(i,\varphi)=d_{1i}$.

$P_k(i,S)$ 为最优决策函数，它表示从 1 城开始经 k 个中间城市的 S 集到 i 城的最短路线上紧挨着 i 城前面的那个城市.

对于四个城市旅行推销员问题，城市间的距离矩阵如表 3-7 所示：

表 3-7

距离	1	2	3	4
1	0	8	5	6
2	8	0	8	5
3	5	8	0	5
4	6	5	5	0

由边界条件可知

$f_0(2,\varphi)=d_{12}=8$，$f_0(3,\varphi)=d_{13}=5$，$f_0(4,\varphi)=d_{14}=6$.

当 $k=1$ 时，即从 1 城开始，中间经过一个城市到达 i 城的最短距离是：

经 3 城到达 2 城　$f_1(2,\{3\})=f_0(3,\varphi)+d_{32}=5+8=13$

经 4 城到达 2 城　$f_1(2,\{4\})=f_0(4,\varphi)+d_{42}=6+5=11$

经 2 城到达 3 城　$f_1(3,\{2\})=f_0(2,\varphi)+d_{23}=8+8=16$

经 4 城到达 3 城　$f_1(3,\{4\})=f_0(4,\varphi)+d_{43}=6+5=11$

经 2 城到达 4 城　$f_1(4,\{2\})=f_0(2,\varphi)+d_{24}=8+5=13$

经 3 城到达 4 城　$f_0(4,\{3\})=f_0(3,\varphi)+d_{34}=5+5=10$

当 $k=2$ 时，即从 1 城开始，中间经过两个城市（它们的顺序任意）到达 i 城的最短距离是：

经 $\{3,4\}$ 到达 2 城：$f_2(2,\{3,4\})=\min[f_1(3,\{4\})+d_{32}$,

$f_1(4,\{3\})+d_{42}]=\min[11+8,10+5]=15$

$\therefore P_2(2,\{3,4\})=4$，即从 1 城开始经两个中间城市$\{3,4\}$，到 2 城的路线上紧挨 2 城的为 4，即 $1\to3\to4\to2$.

经$\{2,4\}$到达 3：$f_2(3,\{2,4\})=\min[f_1(2,\{4\})+d_{23},$

$f_1(4,\{2\})+d_{43}]=\min[11+8,13+5]=18$

$\therefore P_2(3,\{2,4\})=4$，即为 $1\to2\to4\to3$.

经$\{2,3\}$到达 4：$f_2(4,\{2,3\})=\min[f_1(2,\{3\})+d_{24},$

$f_1(3,\{2\})+d_{34}]=\min[13+5,15+5]=18$

$\therefore P_2(4,\{2,3\})=2$，路线为 $1\to3\to2\to4$.

当 $k=3$ 时，即从 1 开始，中间经三个城市(顺序任意)回到 1 城的最短距离是：

$f_3(1,\{2,3,4\})=\min[f_2(2,\{3,4\})+d_{21},f_2(3,\{2,4\})+d_{31},$

$f_2(4,\{2,3\})+d_{41}]=\min[15+8,18+5,18+6]=23$

$\therefore P_3(1,\{2,3,4\})=2$ 或 3. 即到达 1 城前紧挨着 1 的城市应是 2 或 3.

由此可推知，推销员的最短旅行路线是 $1\to3\to4\to2\to1$ 或 $1\to2\to4\to3\to1$，最短距离为 23.

实际中很多问题都可以归结为旅行商这类问题. 如物质运输路线中，汽车应走怎样的路线使路程最短；城市里在一些地方铺设管道，管子应走怎样的路线使管子耗费最少等等.

例 3-12 同顺序流水作业任务安排问题.

设有 m 台机器$(M_1,M_2,\cdots,M_m)$，n 项任务$(J_1,J_2,\cdots,J_n)$，加工顺序相同，依次为 $M_1,M_2,\cdots,M_m$.

$$T=(t_{ij})_{m\times n}$$

为加工时间矩阵，即 t_{ij}为任务 J_j 在机器 M_i 上加工所需的时数.

下面仅就 $m=2$ 的情形加以讨论，令

$$S_0=\{J_1,J_2,\cdots,J_n\},N=\{1,2,\cdots,n\}.$$

n 个任务的加工顺序不同，从第一个任务在 M_1 机器上加工开始，到最后一个任务在 M_2 机器上加工完毕为止，所需的时间可能

各不相同. 最佳的安排要使得 M_2 机器的空闲时间达到最少，而机器 M_1 没有这个问题.

设 S 是任务的集合，若机器 M_1 开始加工 S 中的任务时，M_2 机器在加工其他任务，t 时刻后才可利用. 在这样的条件下，加工 S 中任务所需的最短时间设为 $T(S;t)$. 则有

$$T(S;t)=\min_{J_i\in S}\{t_{1i}+T(S\setminus\{J_i\};t_{2i}+\max\{t-t_{1i},0\})\}$$

其中 $t_{2i}+\max\{t-t_{1i},0\}$，可从图 3 8 中得出.

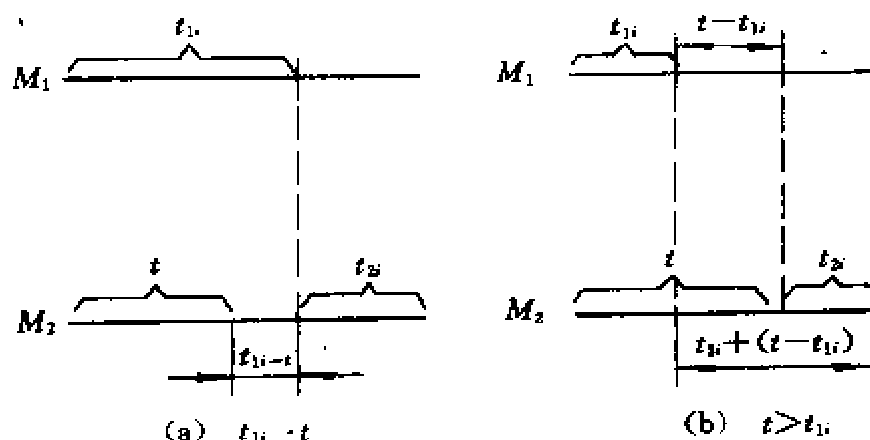


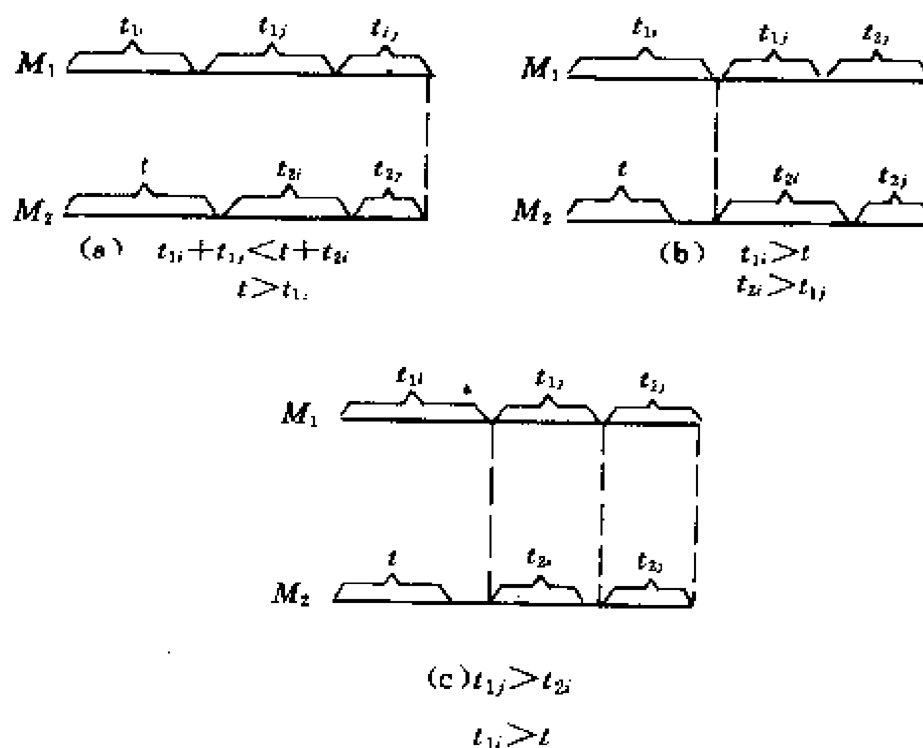
图 3-8

如果最佳方案是 J_i 在前，J_j 在后，则

$$\begin{aligned}T(S;t)&=t_{1i}+T(S\setminus\{J_i\};t_{2i}+\max\{t-t_{1i},0\})\\&=t_{1i}+t_{1j}+T(S\setminus\{J_i,J_j\};t_{2j}+\max\{t_{2i}\\&\quad+\max\{t-t_{1i},0\}-t_{1j},0\})\\&=t_{1i}+t_{1j}+T(S\setminus\{J_i,J_j\},T_{ij})\end{aligned}$$

$$\begin{aligned}T_{ij}&=t_{2j}+\max\{t_{2i}+\max\{t-t_{1i},0\}-t_{1j},0\}\\&=t_{2i}+t_{2j}-t_{1j}+\max\{\max\{t-t_{1i},0\},t_{1j}-t_{2i}\}\\&=t_{2i}+t_{2j}-t_{1j}+\max\{t-t_{1i},t_{1j}-t_{2i},0\}\\&=t_{2i}+t_{2j}-t_{1i}-t_{1j}+\max\{t,t_{1i},t_{1i}+t_{1j}-t_{2i}\}\\&=\begin{cases}t+t_{2i}+t_{2j}-t_{1i}-t_{1j}, & \text{若 } \max\{t,t_{1i},t_{1i}+t_{1j}-t_{2i}\}=t\\ t_{2i}-t_{2j}-t_{1j}, & \text{若 } \max\{t,t_{1i},t_{1i}+t_{1j}-t_{2i}\}=t_{1i}\\ t_{2j}, & \text{若 } \max\{t,t_{1i},t_{1i}+t_{1j}-t_{2i}\}=t_{1i}+t_{1j}+t_{2i}\end{cases}\end{aligned}$$

参见图 3-9.



(a) $t_{1i}+t_{1j}<t+t_{2i}$

$t>t_{1i}$

(b) $t_{1i}>t$

$t_{2i}>t_{1j}$

(c) $t_{1j}>t_{2i}$

$t_{1i}>t$

图 3-9

如果最优次序 $J_i\to J_j$ 的加工顺序互换，则

$$\overline{T}(S;t)=t_{1i}+t_{1j}+T(S\setminus\{J_i,J_j\};T_{ji}),$$

其中，

$$T_{ji}=t_{2i}+t_{2j}-t_{1i}-t_{1j}+\max\{t,t_{1j},t_{1i}-t_{1j}-t_{2j}\},$$

如果

$$\max\{t,t_{1i}+t_{1j}-t_{2i},t_{1i}\}\leqslant\max\{t,t_{1i}+t_{1j}-t_{2j},t_{1j}\} \tag{1}$$

则

$$T(S,t)\leqslant\overline{T}(S;t)$$

进一步可看到，如果

$$t_{1i}+t_{1j}+\max\{-t_{2i},-t_{1j}\}\leqslant t_{1i}+t_{1j}+\max\{-t_{2j},-t_{1i}\},$$

即
$$\min\{t_{2j},t_{1i}\}\leqslant\min\{t_{2i},t_{1j}\} \tag{2}$$

则(1)式成立.

(2)式就是著名的 Johnson 公式，即若(2)式成立，则任务 J_i 安排在 J_j 之前加工．这说明在 M_1 上加工的时间短的任务应优先，而在机器 M_2 上加工时间短的任务应排在后面．因而令 t_{11},t_{12}，

$t_{21}, t_{22}, \cdots, t_{n1}, t_{n2}$按从小到大的顺序排列，若最小的是 t_{k1}，则 J_k 排在第一个，若 t_{k2} 为最小，则 J_k 排在最后一个，并从序列中去掉 t_{k1} 和 t_{k2}，然后再依次观察余下的序数中最小数.

现在某印刷厂有 6 项加工任务，对印刷车间和装订车间所需时间见表 3-8.

表 3-8　　　（时间单位：天）

车间＼任务	J_1	J_2	J_3	J_4	J_5	J_6
印刷车间	3	12	5	2	9	11
装订车间	8	10	9	6	3	1

对矩阵 $T=(t_{ij})_{2\times 6}$的元素从小到大按次排序得

$$\begin{array}{cccccccccccccccccccccccc} 1 & < & 2 & < & 3 & \leqslant & 3 & < & 5 & < & 6 & < & 8 & < & 9 & \leqslant & 9 & < & 10 & < & 11 & < & 12 \\ \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow \\ t_{26} & & t_{14} & & t_{11} & & t_{25} & & t_{13} & & t_{24} & & t_{21} & & t_{15} & & t_{23} & & t_{22} & & t_{16} & & t_{12} \end{array}$$

故最佳的加工顺序应为

$$J_4 \to J_1 \to J_3 \to J_2 \to J_5 \to J_6$$

加工总时数为 43 天.

二、NP 完全问题

离散问题的求解常常要从有限个方案中选出一个满意的结果来. 也许有人认为，从有限个方案中挑选一个，总是比较容易的. 然而，事实并非如此，关键在于问题的规模.

由于计算机的出现，人们对问题的求解在观念上发生了改变. 一个在理论上可解的问题如果在求解时需要花费相当多，以至于太合理的时间（如几百年甚至更长时间），我们不能认为它已解决，而应当努力寻找更好的算法.

如何比较算法的好坏呢？从不同的角度出发可以有不同的回答. 这里，仅就算法的计算速度作一个十分粗略的比较.

设有一台每小时能进行 M 次运算的计算机. 并设问题已有两种不同的算法，算法 A 对规模为 n 的问题约需作 n^2 次运算，算法 B 则约需作 2^n 次运算. 运用算法 A 在一小时内大约可解一个规模为 $\sqrt{M}$ 的问题，而算法 B 则大约可解一个规模为 $\log_2 M$ 的问题.

现在假设计算机有了改进，例如计算速度提高了 100 倍. 此时，利用算法 A 能求解的问题规模增大了 10 倍，利用算法 B 可解的问题规模只增加了 $\log_2 100<7$. 前者得到了成倍的增加，而后者则几乎没有什么改变，今天无法求解的问题，将来也很少有希望解决. 由于这一原因，对算法作如下分类.

定义 1(多项式算法)　设 A 是求解某类问题 D 的一个算法，n 为问题 D 的规模，用 $f_A(D,n)$ 表示用算法 A 在计算机上求解这一问题时需作的初等运算的次数. 若存在一个多项式 $P(n)$ 和正整数 N，当 $n\geqslant N$ 时，总有 $f_A(D,n)\leqslant P(n)$(不论求解的 D 是怎样的具体实例)，则称算法 A 是求解问题 D 的一个多项式算法.

定义 2(指数算法)　设算法 B 是求解某类问题 D 的一个算法，若存在一个常数 $k>0$，对任意 n，总可以找到问题 D 的一个规模为 n 的实例，用算法 B 求解时，所需的计算量约为 $f_B(D,n)=O(2^{kn})$，则称 B 为求解问题 D 的一个指数算法.

多项式算法被称为是“好”算法(或有效算法)，而指数算法则一般认为是“坏”算法，因为它只能用来求解规模很小的问题.

这样看来，对一个问题只有在找到求解它的多项式算法后才能较为放心. 然而十发可惜的是，对于许多具有十分广泛应用价值的离散模型，人们至今仍未找到多项式算法. 现在的任何算法在最坏的情况下计算量均可达到或接近 2^n.

1971 年，S. Cook 发表了“The Complexity of Theorem Proving Procedures”这篇论文. 1972 年，R. Karp 发表了“Reducibility Among Combinationcial Problems”，从此奠定了 NP 完全理论基础. NP 是 Nondeterministic Polgnomial 的编写，即为非确定型的多项式算法. Cook 指出，NP 完全类问题，具有两个性质：

1°这类问题中的任何一个问题至今均未发现有多项式算法.

2°只要其中任一个问题找到了多项式算法,那么其他所有问题也就有了多项式算法.

现在,NP 完全类中的问题已被扩充到了四千多个,其中包括前面讨论的旅行商问题.对它们的研究使人们越来越相信这样一个猜测:对这类问题也许根本不存在多项式算法.

三、近似算法

当前 NP-完全理论的研究中,有两个问题特别引人注目.其一是着眼于证明这一类问题的等价性,即证明它们的困难程度相当.另一个是寻找近似算法.

我们把注意力集中在按照平均性态而不是最坏情况性态的原则来寻找较好的算法,或者更实际地说,寻找一个对于常见的输入能较好地工作的算法.这种选择依赖于经验胜过依赖于严格的分析.这些算法虽然不能保证是最优解,但却是接近最优解的快速算法,它们被称为近似的或启发式算法.不少近似算法的策略和启发式方法非常简单和直接,但对某些问题来说,却提供了意想不到的好结果.

近似算法中最为常用的可能莫过于贪婪算法(Greedy Algorithm),它们的基本策略是,每个选择对象在当时应使度量最优化.

前面我们讨论过旅行商问题,它的一般提法是,记出发城市为0,拟经商的城市分别记为 $1,2,\cdots,n$. 于是问题化为求一个$\{1,2,\cdots,n\}$的排序$\{i_1,i_2,\cdots,i_n\}$,使$\sum\limits_{k=0}^{k}d(i_k,i_{k-1})$最小,其中 $d(i_k,i_{k+1})$为城市 i_k 到 i_{k+1}的距离,$i_0=i_{n-1}=0$. 这是问题的静态模型或组合模型.显然,$\{1,2,\cdots,n\}$的不同排序为 $0(n!)$个,当 n 稍大时,根本无法找出最优方案.例如要想找出遍访美国五十个洲府的最优方案,按照一一对比的穷举方法,需要在计算机上运算数十亿年(对

这个特殊问题，现在已找到好一些的算法）.

例 3-13 旅行商问题的近似解法.

旅行商问题在§6.2已利用动态规划求解（当规模很小时，$n=4$）. 这里讨论一种贪婪算法——最近邻法. 设 n 个城市为 $C=\{C_1,C_2,\cdots,C_n\}$，旅行的出发点设在 C_1，下一个城市应选取与 C_1 点距离最近的点.

一般地，若城市 C_j 简记为 j，假设前面的旅行路线为

$$j_1\to j_2\to\cdots\to j_k \qquad k<n$$

下一个城市 j_{k+1} 应选取其余尚未通过的 $n-k$ 个城市中离 j_k 最近的一个城市.

从 j_1 出发，j_2 的确定要在 j_1 以外 $n-1$ 个城市中进行选择，求其中与 j_1 距离最短者要作 $n-2$ 次比较；j_3 的选择要进行 $n-3$ 次比较；依此类推，故最近邻法共需比较次数为

$$(n-2)+(n-3)+\cdots+1=\frac{1}{2}(n-1)(n-2)$$

最好的情况下，最近邻法可能得到最佳的路径.

令 T_n^* 表示 n 个城市的旅行商问题的最优回路，N_n 为用最近邻法所得的近似解的路径.

假定 $D=(d_{ij})_{n\times n}$ 为 n 个城市间的距离矩阵，$n\geqslant 2$，并满足 1° $d_{ij}=d_{ji}$；2° $d_{ij}\leqslant d_{ik}+d_{kj}$.

设用最近邻法得到的 n 条边的长度为

$$l_1\geqslant l_2\geqslant l_3\geqslant\cdots\geqslant l_n,$$

立即有
$$|T_n^*|\geqslant 2l_1$$

定理 1° $|T_n^*|\geqslant 2\sum_{i=k+1}^{2k} l_i,\ 1\leqslant k\leqslant\left[\frac{n}{2}\right]$.

2° $2N_n\leqslant(T\log_2 n7+1)T_n^*$

按最近邻法，以 1 为出发点，§6.2 中旅行商问题的近似解为 $1\to3\to4\to2\to1$，达到最优解. 以 1 为终点，以倒推方式采用最近邻法，得问题的近似解为 $1\to2\to4\to3\to1$，也达到最优解.

旅行商问题的另一个近似算法是最近插入法. 其基本思想是对由 n 个点中的某 k 点构成的一个最佳回路 T_k,陆续加上一个和 T_k 上的点最接近的点组成新的最佳回路 T_{k+1},开始时可任取一个顶点,直到最后包含 n 个顶点的 T_n 为止.

确定哪个城市加入到 T_k 中来?加入到什么地方?这比最近邻法要复杂,但可在$O(h^2)$时间内完成,即最近插入法的复杂性为$O(h^2)$. 设最近插入法所得的旅行回路的长度为 I_n,则有

$$I_n \Big/ I_n^* < 2$$

对于 § 6.2 中的四城市旅行商问题,求解过程如下:

从 1 出发. 与 1 最近的点为 3,故构成 $T_2=\{1,3\}$. 考察点 2 和 4:点 2 到 T_2 中各点的距离分别为 $d_{21}=8$,$d_{23}=8$,$d_{21}+d_{23}=16$;点 4 到 T_2 中各点的距离分别为 $d_{41}=6$,$d_{43}=5$,$d_{41}+d_{43}=11$,故将点 4 加入到 T_2,构成 T_3. 为确定点 4 加入到什么地方,比较 1→4→3→1 和 1→3→4→1,路径长分别为 6+5+5=16 和 5+5+6=16,路径长相同,任取一种为 T_3 回路. 例如 1→3→4→1.

考虑将点 2 加入 T_3 构成 T_4,加入后可能形成的回路为 1→2→3→4→1(路径长 27),1→3→4→2→1(路径长 23)和 1→3→2→4→1(路径长 24),T_4 的最佳回路为 1→3→4→2→1. 达到最优解.

例 3-14 装箱问题.

设有一批容积为 1 的箱子和 n 个物体 S_i,$i=1,2,\cdots,n$,每个 S_i 的体积也用 S_i 表示,$0<S_i\leqslant 1$. 现在的问题是如何设法将 S_1,S_2,…,S_n 放入尽可能少的箱子中.

若考察所有将 n 项之集合分划成 n 个或少于 n 个的子集(穷举法),最优解就可以找到,但这种允许的分划总数超过$\left(\frac{n}{2}\right)^{n/2}$. 以下述实例来说明两种近似算法的特点.

现有 9 个物体,其大小分别为 0.2,0.2,0.7,0.8,0.3,0.6,0.3,0.2,0.6,试将它们放在尽可能少的单位容积的箱子中.

近似方法一:由大到小装填法

将物体由大到小排成非增序列，每个物体依次放在它第一个放得进的箱子中.

将 9 个物体按大小排序为 0.8，0.7，0.6，0.6，0.3，0.3，0.2，0.2，0.2，记为 $i_1,i_2,\cdots,i_9$. 显然，按此算法，0.8，0.7，0.6，0.6 分别占据一个箱子. 箱子编号为 1，2，3，4. $i_5=0.3$，能装下的第一个箱子为 2 号；$i_6=0.3$，能装下的第一个箱子为 3 号；$i_7=0.2$，能装下的第一个箱子为 1 号；$i_8=i_9=0.2$，分别搜索后，都装入 4 号，故装箱结果为

0.8，0.2	0.7，0.3	0.6，0.3	0.6，0.2，0.2

4 个箱子即可装完，达到最优解.

近似方法二：最佳存放法

将物体由大到小排成非增序列，在存放过程中，尺寸为 S 的物体放在箱子 B_j 中仅当在该物体足以存放的诸箱子中 B_j 是最满的一个，也就是说，若 b_j 是存放在箱子 B_j 中物体的尺寸，则 b_j 是服从要求 $b_j+S\leqslant 1$ 中最大的一个.

在本题中，$b_1=0.8$，$b_2=0.7$，$b_3=0.6$，$b_4=0.6$，分别放入箱子 B_1，B_2，B_3，B_4 中. $i_5=0.3$，将 i_5 放入后最满的箱子是 B_2，故 i_5 放入 B_2；$i_6=0.3$，放入 i_6 后最满的箱子是 B_3，故 i_6 放入 B_3；$i_7=0.2$，放入 i_7 后最满的箱子是 B_1；$i_8=0.2$，放入 i_8 后最满的箱子是 B_4，将 i_8 放入 B_4；$i_9=0.2$，放入 i_9 后最满的箱子是 B_4，将 i_9 放入 B_4，这样得到和方法一相同的结果，也达到最优解.

然而，一般近似算法并非对所有问题都这么幸运. 而且近似算法的理论分析往往比较困难.

例 3-15 任务安排问题.

设有 l 台完全相同的机床：$m_1,m_2,\cdots,m_l$；加工 n 个彼此无关的任务：$j_1,j_2,\cdots,j_n$；所需的时间分别为 $t_1,t_2,\cdots,t_n$. 问如何安排加工顺序，使最后全部结束的时间最短.

l 台相同的机器的任务安排. 当 $l=2$ 时实际上就是划分问题.

设完成任务 j_i 所需时间为 t_i，$1\leqslant i\leqslant n$，任务安排问题相当于对集合

$$A=\{t_1,t_2,\cdots,t_n\}$$

的划分. $l>2$ 是它的自然推广.

近似算法 1.

对时间序列进行排序. 不失一般性，假定

$$t_1\geqslant t_2\geqslant\cdots\geqslant t_n.$$

这个顺序也就是加工的先后顺序. 即当某台机器空闲时，立即加工剩下需要加工的时间最长的任务. 例如，设 $l=3$，$t_1=10$，$t_2=9$，$t_3=8$，$t_4=7$，$t_5=6$，$t_6=5$，加工顺序为

m_1：$j_1(10)$，$j_5(5)$

m_2：$j_2(9)$，$j_5(6)$

m_3：$j_3(8)$，$j_4(7)$

整体加工时间为 15，达到最优.

设最优的任务安排使完成 n 项任务所需的时间为 T_n^*，近似算法所得的任务安排完成时间为 T_n，则绝对偏差 $\Delta=T_n-T_n^*$，相对偏差 $\delta=\Delta/T_n^*$. 以 A_1 记上述近似算法，则近似算法 A_1 的相对偏差满足

$$\delta\leqslant\frac{1}{3}-\frac{1}{3l}.$$

近似算法 2.

假定已经排序得 $t_1\geqslant t_2\geqslant\cdots\geqslant t_n$. 确定整数 k，先对前 k 个任务求最佳的安排，然后对后 $n-k$ 个任务应用算法 A_1.

例 3-16 $l=2$，$n=10$，$k=4$.

$t_1=15$，$t_2=13$，$t_3=12$，$t_4=10$，$t_5=8$

$t_6=7$，$t_7=6$，$t_8=4$，$t_9=3$，$t_{10}=2$

这个例 3-14 的最佳安排为

m_1：$j_1(15)$，$j_2(13)$，$j_3(12)$

m_2：$j_4(10)$，$j_5(8)$，$j_6(7)$，$j_7(6)$，$j_8(4)$，$j_9(3)$，$j_{10}(2)$

最短时间为 40.

按算法 A_2，第一步对前 4 个任务求最佳安排：

$m_1: j_1(15), j_4(10)$

$m_2: j_2(13), j_3(12)$

最短时间为 25.

第二步按 A_1 近似算法进行如下：

$m_1: j_1(15), j_4(10), j_5(8), j_8(4), j_9(3)$

$m_2: j_2(13), j_3(12), j_6(7), j_7(6), j_{10}(2)$

最短时间也是 40，达到最优解.

对近似算法 A_2 有下述估计：

算法 A_2 的相对偏差 δ 满足

$$\delta \leqslant \frac{1-1/l}{1+[k/m]}.$$

§3.5 图论方法

图论方法是建立离散模型的重要方法，这里选用纽约市街道清扫问题作为实例，说明图论方法在人类实践活动甚至日常生活中大有用武之地. 虽然这一例子与我国目前情况的差距较大，但考虑问题的方法还是值得借鉴的.

例 3-17 纽约市街道清扫问题.

纽约市卫生事业的年预算费用约为两亿美元，其中 1000 万美元用于清扫街道. 政府希望能建立一个数学模型，对清扫路线进行规划，使整个清扫工作费用最低.

该模型涉及到图论和线性规划的知识，由于略去了有关的理论证明，这些知识变得相当直观而易于接受.

一、问题的图示

显然，街区可以用图来表示. 在图论中，图由一些点及一些点

之间的联线构成. 这些点称为顶点. 两点间不带箭头的联线称为边，相应的图称为无向图或简称为图；带箭头的联线称为弧，相应的图称为有向图. 表示街区的图，由于单行街道的存在，通常是有向图. 例如图 3-10 所示就是一个简单的街区图，中路为单行主干线，车辆必须按箭头方向通过.

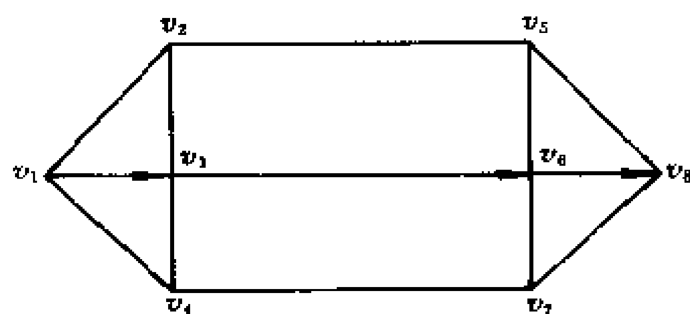


图 3-10 街区图示

二、问题分析

初看起来，问题似乎十分简单，只需把街道分成若干片，每片都能在一个清扫周期内打扫完毕即可. 但事情并非如此. 清扫工作是整个市政管理的一部分，它不是单独存在的，要受到多方面的制约. 这些制约主要是：

第一，像纽约这样的大城市，清扫必须和交通规则相一致. 例如，在车辆繁忙的时段内一些主要街道是不允许清扫的.

第二，清扫街道时路边必须没有停车，即清扫还必须和停车规则一致. 一些小城市可以要求市民服从和清扫相一致的停车规则. 但像纽约这样的大城市则行不通.

第三，城市一般都分成几个行政区，清扫范围可能还不允许越出本区范围.

可见，由于这些限制条件的存在，问题变得复杂起来了. 注意到，在清扫问题中要解决的是如何确定清扫路线，使整个清扫工作费用最低. 从图 3-10 看到，下一步清扫哪一条街道可以看作是在顶点(街角)处决定的. 因此顶点在问题中起着十分重要的作用.

可以设想整个工作是这样进行的.

第一步，根据交通规则和停车规则列出一段时间内可以清扫的街道来.

第二步，找出一个能覆盖这些街道的圈. 在图论中，给定一个图 $G=(V,E)$，其中 V、E 分别是 G 的点集合和边集合. 对一个点边的交错序列 $(V_{i1}, e_{i1}, V_{i2}, e_{i2}, \cdots, V_{ik-1}, e_{ik-1}, V_{ik})$，若 e_{il} 是联接 V_{il} 和 V_{il-1} 的边，则称此序列为连结 V_{i1} 和 V_{ik} 的一条链. 又若 $V_{i1}=V_{ik}$，则称此序列为一个圈.

第三步，将这个圈合理地分成若干段，由清洁工人按规定清扫. 这里所谓的“合理”，指的是既要能够在限定时间内扫完，又要是比较经济的. 这种分段处理的方法，将原问题分割成很多子问题，即在一段规定时间内清扫某些指定街道的问题. 这些子问题还应当通过一些约束条件衔接起来，以保证在一个清扫周期内每条街道被清扫一次且只清扫一次.

这里为简单起见，不准备涉及到子问题的衔接，只着重研究处理子问题的方法.

三、子问题的分析

在按交通规则和停车规则找到某一小段应清扫的街道后，就要寻找最佳路线，这些路线形成一个圈，使这些街道都能被清扫且只被清扫一遍. 但是一般来说，能包含图中所有边的圈通过某些边可能需要两次甚至多次，要寻找的是经过所有顶点走过每条边正好一次的圈（称为欧拉圈）. 如果这种圈存在，当然是最短的. 如果这种圈不存在，就要求重复走过的边的总长为最少的圈. 这个工作就是寻找最小覆盖圈或近似最小覆盖圈.

如何判断子图中是否存在欧拉圈呢？有如下定理.

定理 1 有向图具有欧拉圈的充分必要条件是：(1)此图是连通的；(2)对于每一顶点，内次（进入此顶点的边数）等于外次（离开此顶点的边数）.

这里所说的“连通”，指的是图中任何两个点之间，至少有一条

链.

定理的成立是十分明显的. 因为当且仅当对每一顶点的进入边数等于离开边数时，才能绕行一圈后回到原处. 这是欧拉在研究七桥问题后提出的，它是历史上第一个图论定理. 利用这一定理，可以判断所考察的子图中是否存在欧拉圈. 若存在，设法找出它，即为所求. 遗憾的是，通常是不存在的.

四、子问题的重新叙述

若考察的子图中不存在欧拉圈，则在清扫中势必存在这样的街道，它已经清扫过，而清扫车必须再次通过它. 经调查，当清扫车通过一条不需要清扫的街道时，可以提起扫把，以两倍于清扫时的速度通过它（这段时间称为提升时间）. 由于可以在顶点（街角）处决定下一步清扫哪一条街道，分析这些顶点，可以看出，在那些内、外次不等的顶点处，清扫工人在相应的街角处，一定会遇到提升扫把驶向另一街角重新开始清扫的情况. 由于要清扫的街道是预先指定好的，因而问题的实质就是要找出一个圈，使浪费掉的提升时间总和为最少.

如何来构造一个能覆盖图中所有边的最小圈呢？需要仔细研究哪些内、外次不等的顶点.

记 $d(x)$ 为顶点 x 的外次与内次之差，称 $d(x)$ 为次，即 $d(x)=x$ 点的外次 $-$ x 点的内次. 若 $d(x)<0$，则称 x 为负顶点，若 $d(x)>0$，则称 x 为正顶点. 对负顶点，存在着过剩的进入边；对正顶点，存在着过剩的发出边，过剩边的条数均为 $|d(x)|$.

为了构造出一个覆盖图，必须添加一些新的边（也可以是重复边），使得添加后的新图中每一顶点 x 均满足 $d(x)=0$. 需要特别说明的是，由于添加边是不必清扫的，它们可以不在此子图中，只要在城市所有街道对应的大图中存在即可. 当然希望添加的边的总长最小. 这样，作下述定义.

定义 设 G 是一个每边都附有一个长度的有向图，H 是包含

G 的(大)图，其边也附有长度. 若 A 是 H 中的边的子集且满足：

(1)将 A 添入 G 中作成新图 G'，对 G' 的每一顶点 x，有 $d(x)=0$.

(2)在满足小的集合中 A 具有最小的总长度，则称 A 为 G 关于 H 的最小添加集.

根据前面的分析，有下面定理成立.

定理 G 关于 H 的最小添加集必可分成由 G 的负顶点到正顶点的通路. 若 $d(x)=-k$(或 $+k$)，则有 k 条这样的通路由 x 处发出(或进入).

下面的例子说明如何应用上述定理.

在图 3-11 中，顶点旁括号内的数字为该顶点的次. 虚线边所组成的集合为 A，其中有些边取自整个城市的街区图 H. 显然 A 中的边可以分成由负顶点到正顶点的通路.

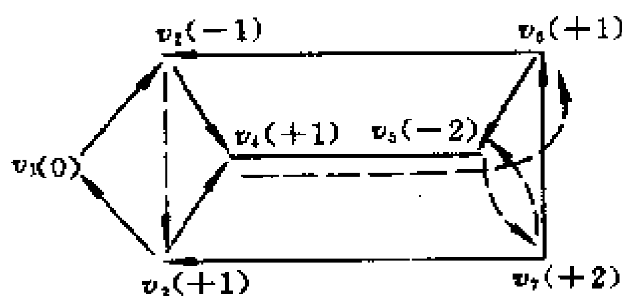


图 3-11 负、正顶点的配对

五、求取最小添加集

现在，问题归结为正负顶点间的配对，即根据 H 的边找出各负正顶点间的最短路. 求图中任意两顶点间的最短路具有专门的有效算法(例如动态规划解法)不在这里讨论. 现在假定已求出了任意一对负正顶点间的最短距离，以 d_{ij} 记负顶点 x_i 到正顶点 x_j 的最短距离，这就构成一个矩阵 $D=(d_{ij})$，在 D 的最右边添加一列，元素为 $b_i=|d(x_i)|$，即 x_i 应发生的边数. 再在 D 的最下面添加一行，元素为 $C_j=d(x_j)$，即 x_j 应收到的边数，构成一个新矩阵

D'.

例如，设图 G 有三个负顶点 x_1，x_2，x_3 和三个正顶点 y_1，y_2，y_3，并设 $d(x_1)=-2$，$d(x_2)=d(x_3)=-1$，$d(y_1)=d(y_2)=1$，$d(y_3)=2$. 假如已求出每对负正顶点间的最短距离，则得到矩阵 D 和 D'：

$$D=\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array}\begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{bmatrix} 3 & 6 & 7 \\ 8 & 2 & 4 \\ 5 & 4 & 1 \end{bmatrix} \end{array}$$

$$D'=\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ \\ \end{array}\begin{array}{c} \begin{array}{cccc} y_1 & y_2 & y_3 & \end{array} \\ \begin{bmatrix} 3 & 6 & 7 & 2 \\ 8 & 2 & 4 & 1 \\ 5 & 4 & 1 & 1 \\ 1 & 1 & 2 & \end{bmatrix} \end{array}$$

求最小添加集 A，是企图在负顶点到正顶点间添加一些路，使得对每一顶点 x，有 $d(x)=0$，且 A 具有最小的总长度. 因此可以把它认为是一个"运输问题"，即将负顶点认为是有货发出的"城市"，正顶点是收货的"城市"，添加一些路，就是确定运输方案，将货运到收货的城市. 对于负顶点，$b_i=|d(x_i)|$ 可认为是"供应量"，对于正顶点，$c_i=d(x_i)$ 可以认为是"需求量"，这里考虑产销平衡的问题，即总供应量等于总需求量. 这一问题的数学模型可表示为：

$$\min A \text{ 的总长度}=\sum_i\sum_j d_{ij}x_{ij}$$

$$\sum_j x_{ij}=b_i=|d(x_i)| \qquad x_i \text{ 的发出边数}$$

$$\sum_i x_{ij}=c_j=d(x_j) \qquad x_j \text{ 的收到边数}$$

$$x_{ij}\geqslant 0 \text{ 且为整数} \qquad x_i \text{ 到 } x_j \text{ 的边数}$$

因此 D' 可表示为

$$
D' = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ \text{需求量} \end{array}
\begin{array}{cccc} y_1 & y_2 & y_3 & \text{供应量} \\ \end{array}
\left[\begin{array}{cccc} 3 & 6 & 7 & 2 \\ 8 & 2 & 4 & 1 \\ 5 & 4 & 1 & 1 \\ 1 & 1 & 2 & 4 \end{array}\right]
$$

右下角的 4 既是总供应量又是总需求量.

运输问题是线性规划的一个特例. 求解上述运输问题，可利用康脱洛维奇的表上作业法. 其基本步骤为

第一步：找出初始基可行解.

第二步：判别是否达到最优解. 如果已是最优解，则停止，否则转第三步.

第三步：确定换入变量和换出变量，找出新的基可行解. 在表上用闭回路法调整.

第四步：重复第二和第三步，直至得到最优解为止.

在求取初始基可行解时，常用的方法之一是最小元素法. 这方法的基本思想是就近供应，即从距离表中最短距离开始确定供销关系，然后次小，一直到给出初始基可行解为止. 最小元素法是一种贪婪算法.

在上述例子中，计算过程如下：

第一步，在 D' 中找到最短距离为 1（相应为 d_{33}），这表示先在 x_3 到 y_3 间添一路径. 因为 $c_3>b_3$，y_3 除满足上述产销关系外，还有需求. 而 $b_3=1$ 已得到满足，划去 x_3 所在行，改 c_3 为 1，改总需求量为 3，仍称矩阵为 D'. 在表 E 中 e_{33} 的位置写上 1，如图所示：

$$
D' = \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ \text{需求量} \end{array}
\overset{\begin{array}{cccc} y_1 & y_2 & y_3 & \text{供应量} \end{array}}{\left[\begin{array}{cccc} 3 & 6 & 7 & 2 \\ 8 & 2 & 4 & 1 \\ \not{5} & \not{4} & \textcircled{1} & \not{1} \\ 1 & 1 & \not{2}^{1} & \not{4}^{3} \end{array}\right]}
\Rightarrow E = \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}
\overset{\begin{array}{ccc} y_1 & y_2 & y_3 \end{array}}{\left[\begin{array}{ccc} & & \\ & & \\ & & 1 \end{array}\right]}
$$

第二步，在 D' 中找到最短距离 2（相应为 d_{22}），这表示应在 x_2

和 y_2 之间添加一路径. 因为 $c_2=b_2$，表示产销平衡，划去 d_{22}所在行和列，并在 E 中 e_{22}处写上 1. 改总需求量为 2，如图所示

$$
\begin{array}{c} \\ D'= \end{array}
\begin{array}{c} \\ x_1 \\ x_2 \\ \text{需求量} \end{array}
\begin{array}{cccc} y_1 & y_2 & y_3 & \text{供应量} \\ \hline 3 & 6 & 7 & 2 \\ 8 & \textcircled{2} & 4 & 1 \\ 1 & 1 & 1 & 3 \end{array}
\Rightarrow E=
\begin{array}{c|ccc|} & y_1 & y_2 & y_3 \\ x_1 & & & \\ x_2 & & 1 & \\ x_3 & & & 1 \end{array}
$$

第三步，在 D' 中，显然应选 d_{11}和 d_{13}，则产销平衡，在 E 中 e_{11}和 e_{13}处填上 1.

$$
E=\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array}
\begin{array}{ccc} y_1 & y_2 & y_3 \\ 1 & & 1 \\ & 1 & \\ & & 1 \end{array}
$$

则 E 所示即为初始基可行解，经检验这是最优解. 所以运输问题的解为 $x_1\to y_1, x_1\to y_3, x_2\to y_2, x_3\to y_3$，这些最短路径各取一次就得到最小添加集 A.

这一运输问题在例题中的对应关系为

$$x_1\to v_5,\ x_2\to v_4,\ x_3\to v_2,\ y_1\to v_3,\ y_2\to v_6,\ y_3\to v_7.$$

因此，若在图中用虚线表示添加的路线，则如图 3-12 所示.

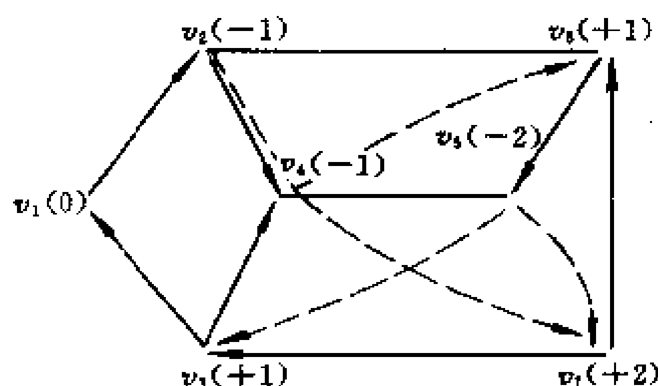


图 3-12　G 的最小添加集

六、实际运行中的若干问题

在求出由 G 扩充成的最小覆盖圈 G' 后，还需为清扫工人指出

一条较好的实际运行路线. 不同的运行方式在工作时的方便程度可以相差很大. 清扫车有时要扫街道左边，有时要扫右边；有时要提升扫把，快速通过不清扫的街道. 在拐角处，有时要左转弯，有时要右转弯，甚至U形转弯，事先规划好是十分有益的. 纽约市是这样处理这一问题的，他们首先引入一个权因子，以反映街角处可能遇到的各种情况的麻烦程度，综合清扫工人的意见定出权因子的值，如表 3-9 所示：

表 3-9　街角处决策的权因子

街角决策	权因子
直走	0
大转弯	4
小转弯	1
U形转弯	8
转换扫把	10
提升扫把	5

进而对每一街角 V_k 找出一个进出边之间的一对搭配，使得权因子总和最小(即操作最方便)，这种搭配关系唯一地确定了实际运行路线.

这样，对于每一进(出)边多于一条的顶点 v_k，得到一个权因子组成的矩阵 $w^{(k)}=(w_{ij})$，其行对应进边，其列对应出边，而 w_{ij} 则为由 i 边进入，由 j 边离开时司机需采取的措施所对应的权因子. 例如，对图 3-13 表示的街角 v_k 可以得出：

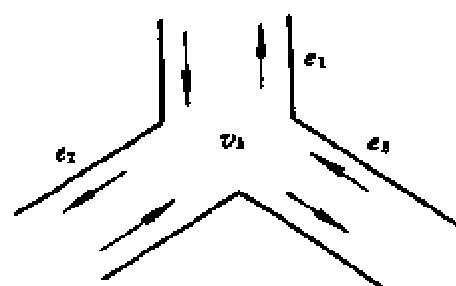


图 3-13　三叉路口 v_k 示意图

$$w^{(k)}=\begin{array}{c} \\ e_1 \\ e_2 \\ e_3 \end{array}\begin{array}{c} \begin{array}{ccc} e_1 & e_2 & e_3 \end{array} \\ \begin{bmatrix} 8 & 1 & 4 \\ 4 & 8 & 1 \\ 1 & 4 & 8 \end{bmatrix} \end{array}$$

其中，e_{ii}表示U形转弯，权因子为 8；e_{12}为小转弯，权因子为 1；e_{13}为大转弯，权因子为 4，依此类推.

在此研究的搭配，实质上是简单的指派问题. 指派问题的一般提法是：某单位需完成 n 项任务，恰好有 n 个人可承担这些任务，由于每个人的专长不同，各人完成任务不同，效率也不同，问应指派哪个人去完成哪项任务，使完成 n 项任务的总效率最高. 其数学模型的一般表达式为

$$\min Z=\sum_{i}\sum_{j}c_{ij}x_{ij}$$

$$\sum_{i}x_{ij}=1,\quad j=1,2,\cdots,n(j\text{ 任务只能由 1 人完成})$$

$$\sum_{j}x_{ij}=1,\quad i=1,2,\cdots,n(i\text{ 任务只能由 1 人完成})$$

$$x_{ij}=1\text{ 或 }0$$

其中，$x_{ij}=1$，表示指派第 i 人完成第 j 项任务，否则 $x_{ij}=0$.

指派问题是 0—1 规划的特例，也是运输问题的特例，用匈牙利算法容易求解. 在这里，最佳搭配是显然的，即 $e_1\rightarrow e_2, e_2\rightarrow e_3, e_3\rightarrow e_1$，在此搭配下，权因子总和达到极小值了. 当然这是比较简单的情况.

至此解决了子问题. 最后来分析一个实例，其中要考虑子问题的连接.

图 3-14 所示是一个街区图，其中箭线为单行线，边上的数字是清扫该街道所需要的时间（提升扫把时，时间减半）. 图 3-15 中的箭线表示上午 8：00—9：00 禁止停车的街道，现在要求充分利用这一小时不停车时间清扫完图 3-15 中的街道. 试给出一个较好的方案.

第一步，找出图 3-15 中的负顶点和正顶点，并对这些顶点求出 $d(x)$，标在图中. 负顶点为 3，7，12，14，16，18，正顶点为 1，4，5，13，15，17，每一顶点的“供应量”或“需求量”均为 1.

第二步，求出每对负、正顶点间的距离，为方便起见，以清扫时

间表示之(纽约市是采用引入坐标的方法,利用计算机计算的). 根据给出的两个图,求得反映负正顶点间距离的矩阵

$$D=\begin{array}{c} \\ 3 \\ 7 \\ 12 \\ 14 \end{array}\begin{array}{c} \begin{array}{cccc} 1 & 4 & 5 & 13 \end{array} \\ \begin{bmatrix} 12 & 8 & 16 & 25 \\ 16 & 12 & 20 & 22 \\ 20 & 26 & 24 & 7 \\ 32 & 28 & 36 & 8 \end{bmatrix} \end{array}$$

这里没有包括负顶点 16、18 和正顶点 15、17. 因为十分明显, 16 应与 15 配对,18 应与 17 配对.

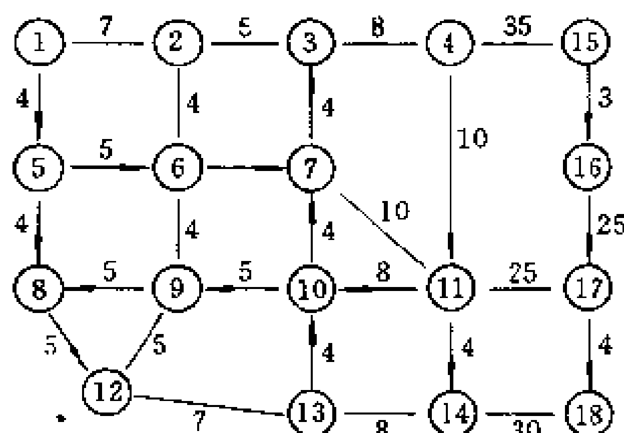


图 3-14　街区图

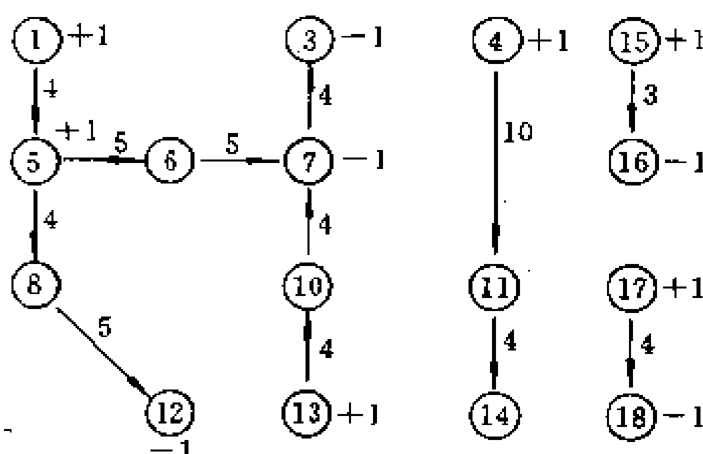


图 3-15　需清扫的街道

求解与 D 相应的运输问题. 由于次数都是 1,问题已退化为指

派问题.得到如下配对:3→4,7→1,12→5,14→13.添入相应的最短路,得到扩充图 G,见图 3-16,增添路总长为 63 分钟,折合成清扫时间为 $31\frac{1}{2}$分钟(在增添路上是将扫把提升起来快速通过的,时间减半).用空箭线表示添加路径.

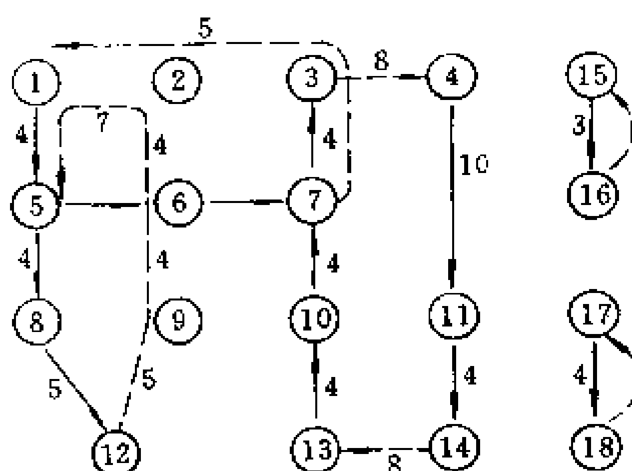


图 3-16 扩充图 G

这里由于进出边配对问题较为简单,不必再对顶点求解指派问题.

图 3-16 中包含三个圈,它们过每条边(包括添加边)正好一次.记左边大圈为 G_1,右边上下两圈分别为 G_2 和 G_3.

第三步,找出三个圈的最短连线.G_1 和 G_2 最接近的两个顶点为 4 和 15,距离为 35 分钟;G_1 和 G_2 最接近的顶点是 11 和 17,距离为 25 分钟;G_2 和 G_3 最接近的顶点是 16 和 17,距离也为 25 分钟.用最近顶点间的最短路将三个小圈连接成一个大圈,见图 3-17,即 $G_1\to G_3\to G_2$.走完整圈需要 $137\frac{1}{2}$分钟,其中 56 分钟是清扫时间,添加路径长 63 分钟,子图连接路径长 100 分钟,故提升时间为 $81\frac{1}{2}$分钟.

第四步,现在作实际运行分配.由于全程需 $137\frac{1}{2}$分钟,因此一台清扫车不可能在一小时内完成,也不可能简单地分成两段 60

分钟的子路，使用两台清扫车同时作业. 当然有很多方式可将全程分成三段，同时使用三台清扫车作业，但这样做并不是最经济的.

注意到从 16→17→11 有一段总长达 50 分钟的提升时间，考虑是否可以利用这一点，即设法将这段安排在分段的开头或结尾，以便在 8：00 以前或 9：00 以后通过它. 设想将图分成两段，以便节省一台车. 然而像图 3-17 这样的圈，这种方法行不通.

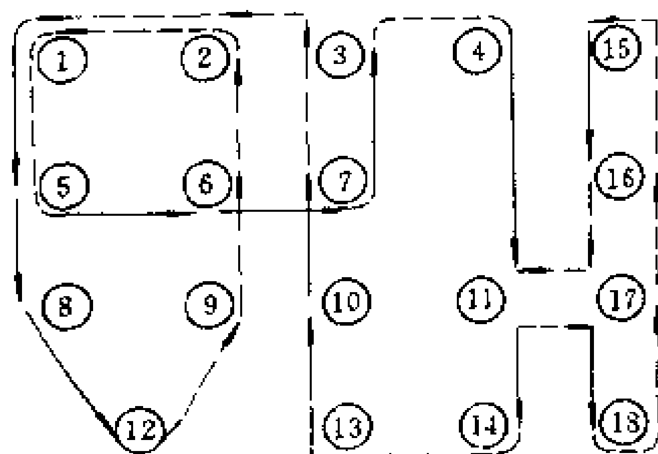


图 3-17　三个小圈连成的大圈

现在改用 18→14 连接 G_1 和 G_3（图 3-18）. 新添路程为 30×2 分钟，表面上看，路程大于用 11→17 连接 G_1 和 G_2，然而得到的新圈可以分成两段，如图 3-19 所示.

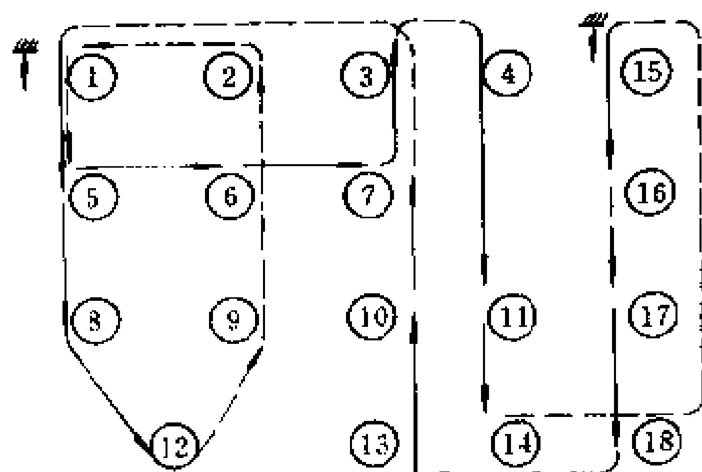


图 3-18　修改后的子图连接

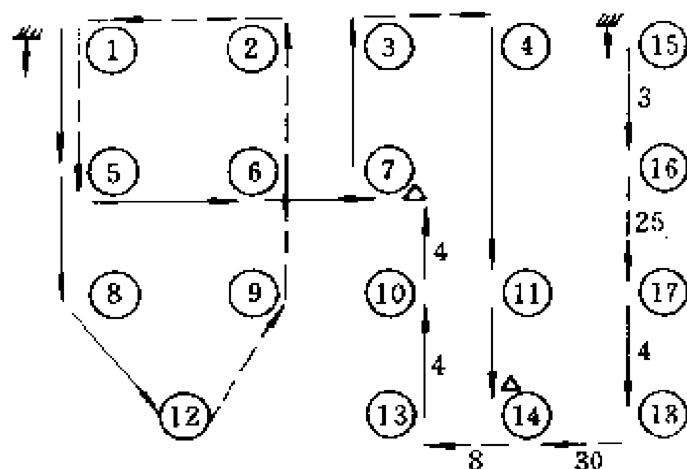


图 3-19　两台车的清扫方案

T_1：1→5→8→12→9→6→2→1→5→6→7→3→4→11→14

T_2：15→16→17→18→14→13→10→7

T_1 需要清扫时间 57 分钟（其中，41 分钟清扫，32 分钟提升时间）. T_2 需要清扫时间 $46\frac{1}{2}$分钟（其中 15 分钟清扫，$31\frac{1}{2}$分钟提升时间. 若将 1→5 安排在第二次通过时清扫，即 T_1 从 5 开始，则 T_1 占用的禁止停车时间又可缩短 2 分钟. 当然清扫车应于 8：00 以前赶到 5 处开始清扫.

现在经过反复筹划，只需派出两辆清扫车即可在 8：00—9：00 内打扫完指定街道. 利用这个模型，估计清扫费用可节省 10%. 而事实上，在哥伦比亚地区采用后，实际节省了 20%以上的费用，经济效益十分可观.

习　　题

1. 在基因遗传过程中，考虑 3 种基因类型：优种 $D(dd)$，混种 $H(dr)$和劣种 $R(rr)$. 对于任意的个体，每次用一混种与之交配，所得后代仍用混种交配，如此继续下去，构造马氏链模型，说明是正则链，求稳态概率及由优种和混种出发的首次返回平均转移次数. 如果改为每次用优种交配，再构造马氏

链模型，说明是吸收链，求由混种和劣种出发变为优种的平均转移次数.

2. 贮物仓库每周或每月清仓一次，检查库存量. 假定货物入库量是确定的，而出库量是随机的，其概率分布由需求情况决定，仓库的容量是有限的. 试建立一个模型分析库存量的变化规律，特别是经过多长时间会出现库满或者库空的情况.

3. 考查水库管理，设入库流量是随机的，概率分布已知，出库流量是确定的，以库容量为状态，构造马氏链模型. 如果入库流量来自上游河流，而河流的流量又与季节有关，问上面的模型应作哪些改变？

4. 一个服务网络由 k 个工作站 $v_1, v_2, \cdots, v_k$ 依次串接而成，当某种服务请求到达工作站 v_i 时，v_i 能够处理的概率为 p_i，转往下一站 v_{i+1} 处理的概率为 $q_i(i=1,2,\cdots,k-1$，设 $q_k=0)$，拒绝处理的概率为 r_i，满足 $p_i+q_i+r_i=1$. 构造马氏链模型，确定到达 v_i 的请求平均经过多少工作站才能获得接受处理或拒绝处理的结果，被接受和拒绝的概率各多大.

5. 在社会系统中常常按照人们的职务或地位划分出许多等级，如大学教师分为教授、讲师和助教，工厂技术人员分为高级工程师、工程师和技术员等等. 不同等级人员的比例形成一个等级结构，合适的，稳定的等级结构有利于教学、科研、生产等各方面工作的顺利进行. 试建立一个模型来描述等级结构的变化状况，根据已知条件和当前的结构预报未来的结构，并讨论为了达到某个理想的等级结构而应采取的策略.

6. 请你帮一位高中毕业生选报高考志愿. 选报时通常要考虑到学校名誉、教学、文体及环境条件，同时又要结合本人兴趣及考试成绩. 在每一因素内还含有子因素，例如教学因素中要考虑到教师水平、学生水平、深造条件等. 考生可填写四个志愿 A、B、C、D. 你如何用层次分析法帮他填志愿.

7. 用层次分析法选择理想交通工具. 不同人外出的目的不同，经济条件不同，体质、心理、经历、兴趣都不同，考虑到安全、舒适、快速、经济、游览等因素，选择何种交通工具(包括飞机、火车、汽车).

8. 学校评选优秀学生或优秀班级，试给出若干准则，构造层次结构模型，可分为相对评价和绝对评价两种情况讨论.

9. 在 Leslie 模型的基础上，建立种群的稳定收获模型，即周期地捕捉，使得每次产量都相同，且在每次收获后剩下的总体年龄分布不变(收获超增长部分)，讨论获得稳定收获的重要条件.

10. 讨论稳定收获的特例①只捕获种群中最年幼的；②随机捕捉.

11. 在饭馆点菜，首先要求包含我们需要的营养成分. 设菜单及包含的营养成分如下表(用 1 和 0 分别表示包含和不包含这种成分). 你如何点菜呢? 如果给这些菜标上价钱，你想在保证营养条件下最省钱，又如何点菜?

表 3-10　菜单及其营养成分

菜单 \ 营养成分	蛋白质	糖	维生素	矿物质
菜肉蛋卷	1	0	1	1
炒猪肝	0	1	0	0
沙　拉	0	0	1	0
红烧排骨	1	0	0	0
咖喱牛肉	0	1	0	0
清汤全鸡	1	0	0	1

12. 若干支球队参加循环比赛，他们两两相互交锋. 假设每场比赛只计胜负，且不允许平局，在循环赛结束后怎样根据他们的比赛成绩排列名次?

13. 一家公司生产若干种化学制品，其中某些制品是互不相容的，如果存放在一起，则可能发生化学反应，引起危险. 因此公司必须把全库分成相互隔离的若干小区，以便把不相容的制品分开存放. 问至少要划分多少小区，存放才能保证安全.

14. 对例 3-1 中的安全渡河问题编写计算机程序求解.

第四章　逻辑方法

逻辑方法是数学理论研究的重要方法.这一方法也可应用于数学建模，即把对象的基本属性抽象成定义、公理，然后运用逻辑推理方法，或者得到满足这些公理的结果，从而提供解决问题的正面答案，或者证明不存在原来意义下的解，而反过来审查对象的属性，以及所作的抽象定义和公理，从而得到对问题及其解决途径的重新认识.逻辑方法适合于社会学和经济学等领域的问题，在决策、对策等学科中得到广泛的应用.

§4.1　实物交换问题

为了能用数学方法对经济问题进行分析，必须对经济社会作出一些假设.基本假设之一是经济社会的所有成员的行为准则都是为了使自己得到的效益尽可能大.由于人们对风险具有不同的态度，对事物也具有不同的倾向或偏好，这些都对决策产生重大影响.因此必须研究人们的偏好关系.假设：

假设 1（唯一性）　若某人对于对象集合中任何两个财货向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^T$ 和 $\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^T$，根据他本人的倾向作出下列三个判断之一：(1)认为 $\boldsymbol{x}$ 比 $\boldsymbol{y}$ 好，记为 $\boldsymbol{x}\textcircled{>}\boldsymbol{y}$；(2)认为 $\boldsymbol{y}$ 比 $\boldsymbol{x}$ 好，记为 $\boldsymbol{y}\textcircled{>}\boldsymbol{x}$；(3)认为 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 无差异，记为 $\boldsymbol{x}\textcircled{=}\boldsymbol{y}$，则称该人在对象集合上有一个偏好关系.

假设 2（单调性）$\forall\ \boldsymbol{x},\boldsymbol{y}$，若 $\boldsymbol{x}\geqslant\boldsymbol{y}$，则 $\boldsymbol{x}\textcircled{$\geqslant$}\boldsymbol{y}$（多多宜善原则）.

假设 3（传递性）　偏好关系满足下列三公理：

$$\boldsymbol{x}\textcircled{$\geqslant$}\boldsymbol{y}\ \text{且}\ \boldsymbol{y}\textcircled{$\geqslant$}\boldsymbol{x}\Rightarrow\boldsymbol{x}\textcircled{=}\boldsymbol{y}$$

$$\boldsymbol{x}\textcircled{$\geqslant$}\boldsymbol{y}\ \text{且}\ \boldsymbol{y}\textcircled{>}\boldsymbol{z}\Rightarrow\boldsymbol{x}\textcircled{>}\boldsymbol{z}$$

$$\boldsymbol{x}\textcircled{>}\boldsymbol{y}\ \text{且}\ \boldsymbol{y}\textcircled{$\geqslant$}\boldsymbol{z}\Rightarrow\boldsymbol{x}\textcircled{>}\boldsymbol{z}.$$

假设 4(平衡性)　若 $\boldsymbol{x}=(x_1,x_2)^T,\boldsymbol{y}=(y_1,y_2)^T$,则对于 $x_1>y_1$,必存在这样的 $y_2>x_2$,使

$$(x_1,x_2)^T \ominus (y_1,y_2)^T$$

这意味着,若消费者减少了第一种财货,则可增加第二种财货来弥补其损失.

假设 5(凸性假设)　若 $\boldsymbol{x}=(x_1,x_2)^T,\boldsymbol{y}=(y_1,y_2)^T$,则 x_1 逐渐减小时,x_2 的增加应越来越大.这反映了消费者的心理状态,当 x_1 越小时,要对 x_2 的补偿越高,且增长很快.

考虑 $\forall\ \boldsymbol{x}$,满足 $\boldsymbol{y} \ominus \boldsymbol{x}$ 的 $\boldsymbol{y}$ 的集合,在 R^2_+ 中形成一条曲线,在经济上称为过 $\boldsymbol{x}$ 的无差异曲线.

由假设 4,无差异曲线为单调降.由假设 5,无差异曲线为下凸,如图 4-1 所示.

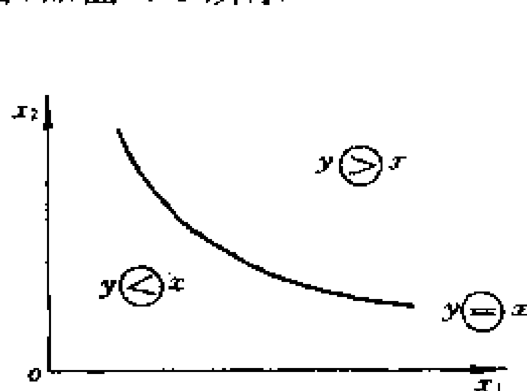


图 4-1

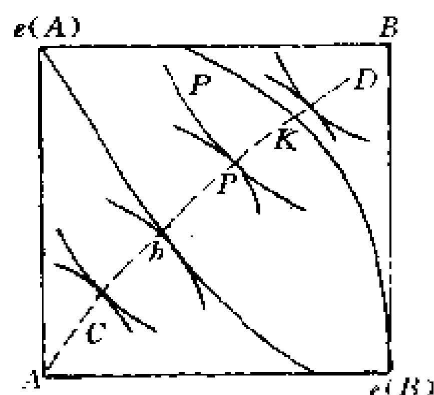


图 4-2　交换问题图示

利用无差异曲线,可以研究市场上的贸易问题.

例 4-1　交换问题.

考虑仅由两个人(A 和 B)和两种产品(例如香烟和馒头)组成的经济系统.如果 A 有 3 盒香烟,B 有 3 斤馒头,A 和 B 把他们持有的产品进行交换,以改善他们现有的消费效用.

由于该系统是封闭的,经过交换后总量并不改变,因此 A、B 各人的产品持有量(即一种分配状态 x_1,x_2)都可用图 4-2 所示矩形中某一点表示这一矩形在经济中称为 Edge worth 盒.交换前的产品配置状态是 $e(A)=(0,3)^T,e(B)=(3,0)^T$.显然 A 有无数条

无差别曲线布满整个矩形，且越往右上方的曲线，满意程度越高. 对 B 也有类似的无差异曲线. 为了得到双方满意的交换方案，将双方的无差异曲线族画在一起，如图 4-2 所示. 这两族曲线的切点连成一条曲线 CD，图中用虚线表示. 可以肯定，双方满意的交换方案都在曲线 CD 上，称之为交换路径. 实际上，假设交换在曲线之外的某一点 P' 进行，又设交换路经 CD 上的 P 点与 P' 点在 A 的同一条无差异曲线上，则对 A 来说，在 P 点和 P' 点进行交换，满意程度是一样的，但对 B 来说，P 点的满意度高于 P' 点，所以交换不可能在 P' 点进行. 即 CD 上的任一点都具有这样的性质：在矩形中再也找不到其他的点，使 A、B 两人中至少有一人获得更多的消费效用，而另一人的消费效用并不减少. 这样的点在经济学中称为 Pareto 最优点. 我们称曲线 CD 为这个经济系统的 Pareto 最优集或合同曲线，也称为第一类最优配置集合.

通过 $e(A)$ 的 A 的无差异曲线和通过 $e(B)$ 的 B 的无差异曲线围成一个区域. 显然在此区域内任何一点的消费效用无论对 A 或 B 而言都比 $e(A)$ 或 $e(B)$ 好. 因此只要这两条曲线围成的区域非空，将原来的分配状态换成区域中任一点，对 A、B 双方都有好处. 注意到，合同曲线 CD 上的 hk 段才是值得考虑的最优配置，称之为这个经济系统的核，也称为第二类最优配置集合，它与初始状态有关.

以上是经济中交换问题的图解法，它提供了一个定性模型.

假设在经济系统中存在一个价格，A 和 B 都接受这个价格，并在此价格下进行交换. 在只有两种商品的经济系统中，实际上只要一个价格就够了，即只要知道馒头和香烟的交换比即可. 记 $\boldsymbol{P}=(P_1,P_2)^T$ 为两种商品的价格，把一盒香烟取作基本价格单位，即 $P_2=1$. 设 $P_1=3$，即 1 斤馒头的价格等于 3 盒香烟的价格.

在上述交换问题中，A 开始有 3 盒香烟，即 $e(\mathrm{A})=(0,3)^T$，因此他的货币收入为

$$P^T \cdot e(\mathrm{A}) = (3,1)\begin{pmatrix}0\\3\end{pmatrix} = 3.$$

若用 $x = (x_1, x_2)^T$ 表示商品向量，则有收入的约束

$$P^T x \leqslant 3 \text{ 或 } 3x_1 + x_2 \leqslant 3$$

如图 4-3 中的阴影三角形. 设他的无差异曲线和直线 $P^T x = 3$ 相切于点 $g(\frac{1}{3}, 2)$，则他应选择使他能获得最大收入的商品组合(或称需求)g.

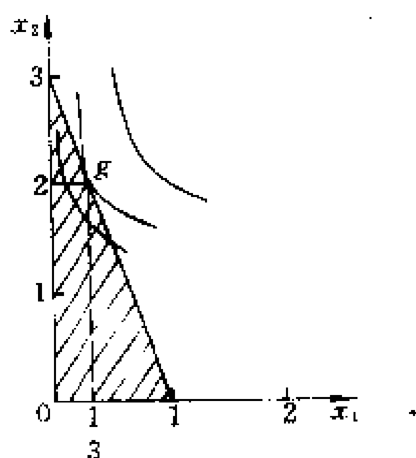


图 4-3　A 的最优交换方案

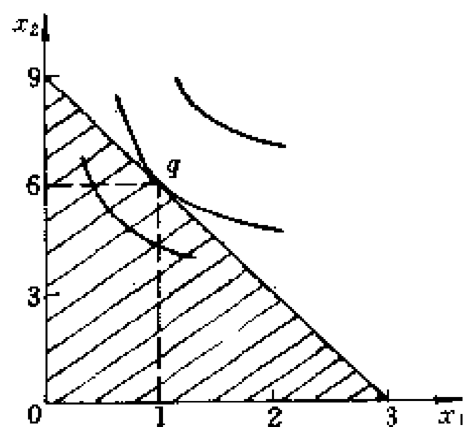


图 4-4　B 的最优交换方案

类似地，B 的开始收入为

$$P^T \cdot e(B) = (3,1)\begin{pmatrix}3\\0\end{pmatrix} = 9$$

其收入约束为 $P^T x \leqslant 9$，　即　$3x_1 + x_2 \leqslant 9$. 设其无差异曲线和收入约束线相切于 $q(1,6)$，如图 4-4 所示.

这样，对于价格 $P = (3,1)^T$，A 在他的收入约束下最理想的需求是$\frac{1}{3}$斤馒头和 2 盒香烟，而 B 的理想消费是 1 斤馒头和 6 盒香烟，总需求为 $1\frac{1}{3}$斤馒头和 8 盒香烟. 这时市场上将会有 $1\frac{2}{3}$斤馒头过剩，而出现 5 盒的短缺. 若要求供需平衡，则必须修改价格. 直观上看来，问题在于香烟价格太低. 有人会问，是否存在这样一个价格 P，使得在此价格下每个人都在各自的收入约束下达到最优，

而且总需求等于总供给？称这种意义下的最优分配为竞争配置.

可以证明在一定条件下竞争配置的存在性，而且交换问题中三种最优配置的关系为

竞争配置$\subset$核$\subset$Pareto 最优点集

而竞争配置集合可以用核的极限来逼近.

求解交换问题至此尚有两个问题没有解决. 一个是如何求取无差异曲线，另一个是 A、B 的最优配置 g 和 q 如何求.

分析无差异曲线，它实质上反映了实物交换问题中，决策者对交换方案的偏好. 对此经济管理学家引进了效用概念作为指标，度量人们对某些事物的主观价值、态度、偏爱和倾向等等. 在风险决策中，用效用这指标来量化决策者对待风险的态度，可以给每个决策者测定他对待风险的态度的效用曲线. 效用值是一个相对指标值，一般可规定：凡对决策者最爱好、最倾向、最愿意的事物（事件）的效用值赋于 1，而最不爱好的赋于 0. 也可以用其他数值范围. 但效用指标无量纲，通过它可以将某些难于量化而有质的差别的事物给于量化.

在第一个问题中，由于无差异曲线实质上是效用函数的等值线 $U(X)=c$，而效用函数可用心理测试的提问法（直接提问和对比提问等）获取，也可以用下列函数拟合

$$U(X)=c_1+a_1(x-c_2)$$

$$U(X)=c_1+a_1[1-e^{a_2(x-c_2)}]$$

$$U(X)=c_1+a_1[2-e^{a_2(x-c_2)}-e^{a_3(x-c_3)}]$$

$$U(X)=c_1+a_1[1-e^{a_2(x-c_2)}]+a_3(x-c_2)$$

$$U(X)=a_1+c_1\cdot a_2[c_1(x-a_3)]^{a_4}$$

$$U(X)=c_1+a_1\log(c_3x-c_2)$$

一旦确定了效用函数后，g 和 q 的求解归结为非线性规划：

$$\begin{aligned}&\max U_A(x)\\ &\text{s.t.}\ P_1x_1+P_2x_2\leqslant \boldsymbol{P}^T\boldsymbol{e}(A)\\ &\qquad x_1,x_2\geqslant 0\end{aligned}$$

和
$$\max U_B(\boldsymbol{x})$$
$$\text{s.t. } P_1x_1+P_2x_2\leqslant \boldsymbol{P}^T\boldsymbol{e}(B)$$
$$x_1, x_2\geqslant 0.$$

§4.2 费用分摊问题

费用分摊(Cost Allocation)问题是综合性建设项目的经济分析中经常遇到的一类问题. 费用分摊问题的最先提出和研究出自于1935年美国对田纳西河流域的综合开发建设项目. 当时的建设费用需要用在防洪、移民、灌溉、国防和能源开发等五个方面，如何在各个参加开发的部门之间分摊投资费用，成为一个难题，分摊得合理与否关系到各部门是否愿意参加合作以及参加后是否具有积极性. 因此，数十年来，分摊问题一直是很多工程技术人员和经济分析人员致力研究的问题，它在经济和社会的其他领域也有广泛的实际背景.

一、传统的分摊方法

1. 一次性分摊

这种方法的基本内容是根据某个数量指标(如设施的使用量、效益等)将总费用按比例一次性分摊. 这样做，虽然计算简单，但在很多场合下并不合理. 例如兴建水库，如果按所需库容大小分摊投资，则防洪部门的分摊偏多，其他部门的分摊可能偏少，矛盾很大.

2. 二次分摊

设有 n 个部门共同参加一个建设项目，记 $I=\{1,2,\cdots,n\}$ 为全体部门的集合. 记总的投资费用为 $C(I)$. 对任意子集 $S\leqslant I$，$C(S)$ 表示团伙 S 单独承接该项目的费用. 将总费用 $C(I)$ 分成两个部分：可分费用和剩余费用. 在目前广泛采用的可分费用剩余效益法(SCRB)中，可分费用定义为

$$m_i=C(I)-C(I\setminus\{i\}),\quad i=1,2,\cdots,n$$

其中，m_i^c 是部门 i 的可分费用，它是第 i 部门加入到项目中后总费用的增加额. 显然，m_i^c 是部门 i 至少应承担的费用. 一般说来，各部门对自己的可分费用额不会持较大的异议.

将可分费用分摊后，剩余费用为

$$E(I)=C(I)-\sum_{i=1}^{n} m_i^c$$

于是每个部门的分摊额 m_i 为

$$m_i=m_i^c+W_iE(I),\quad i=1,2,\cdots,n$$

其中，w_i 为剩余费用分摊比例，满足 $w_i\geqslant 0$，$\sum_{i=1}^{n} w_i=1$. 根据对 w_i 的不同取值可得到不同的二次性分摊方法. 在 SCRB 方法中，取

$$w_i=\frac{\min\{c(i),b(i)\}-m_i^c}{\sum_{i=1}^{n}(\min\{c(i),b(i)\}-m_i^c)}\qquad i=1,2,\cdots,n$$

其中，$c(i)$和 $b(i)$分别为部门 i 参加合作后的获利与单干时的获利.

以上两种传统分摊方法计算简单，要求信息量不大，但合理性较差.

二、对策模型

将合作对策的思想应用于费用分摊问题，起始于 1973 年前后，目前已形成了一类特殊的多人合作对策——费用对策.

1. Shapley 合作对策模型

在合作对策中，合作的目的是为了获得利益，因此，在实际问题中，常常将团体中各种组合的合作获利定义为合作对策的特征函数. 一般，记 n 人集合为 $I=\{1,2,\cdots,n\}$，如果对于 I 的任一子集 $S\subset I$，都对应一个实际函数 $v(S)$，满足

$$v(\varphi)=0$$

$$v(S_1\cup S_2)\geqslant v(S_1)+v(S_2)\quad (S_1\cap S_2=\varphi)$$

则称 v 为定义在 I 上的特征函数. 上述定义中的第二式表示合作

规模扩大时，获利不会减少.

作为参加者，集合 I 中每个成员关心的是自己在合作中应得的收入 $\varphi_i(v)(i\in I)$要合理. Shapley1953 年研究了合作对策中分配的合理性.

设 v 是定义在 $I=\{1,2,\cdots,n\}$上的特征函数，$\varphi(v)=(\varphi_1(v),\cdots,\varphi_n(v))$是互作对策.

公理 1(对称性) 设 π 是 $I=\{1,2,\cdots,n\}$的一个排列，对于 I 的任一子集 $S=\{i_1,i_2,\cdots,i_S\}$，$\pi_S=\{\pi(i_1),\cdots,\pi(i_S)\}$，若再定义一个特征函数，$w(S)=v(\pi S)$，则对于每一个 $i\in I$，有

$$\varphi_i(w)=\varphi_\pi(i)(v)$$

公理 2(有效性) $\sum_{i=1}^{n}\varphi_i(v)=v(I)$

公理 3(无功不受禄) 如果对于所有包含 i 的子集 S，都有 $v(S\backslash\{i\})=v(S)$，则

公理 4(多劳多得) $\varphi_i(v)=0$

若 v' 也是定义在 I 上的特征函数，且 $w=v+v'$，则 $\varphi(w)=\varphi(v)+\varphi(v')$.

Shepleg 首先证明满足公理 1～4 的 $\varphi(v)$是唯一的. 然后构造 $\varphi(v)$为：

$$\varphi_i(v)=\sum_{s\in s_i}w(|S|)[v(S)-v(S\backslash\{i\})]\quad i=1,2,\cdots,n.$$

其中，S_i 是 I 中包含 i 的所有子集，$|S|$是集合 S 中的人数，$w(|S|)$ 是加权因子，由

$$w(|S|)=\frac{(|S|-1)!\ (n-|S|)!}{n!}$$

确定. 式中$[v(S)-v(S\backslash\{i\})]$表示 i 对合作 S 的贡献. $\sum_{S\in S_i}$ 表示对所有包含 i 的集合求和. $\varphi|v)$称为由 v 定义的合作的 Shapley 值. 显然成员 i 的收入是它对各种形式的合作的贡献的加权平均值.

由此，i 在合作中的费用分摊 M_i 为单干时的投资与合作后的

获利之差，即

$$M_i^c = C(i) - \varphi_i(v), i=1,2,\cdots,n.$$

2. 极小费用缺口法(CGA 法)

为改进 SCRB 方法，尽可能合理地确定剩余费用的分摊比例系数 $w_i, i=1,2,\cdots,n$. 记费用缺口(即剩余费用)为 $g^C(I)$，即

$$g^C(I) = C(I) - \sum_{i=1}^{n} m_i^c$$

同样可以定义任何团伙 S 的费用缺口

$$g^C(S) = C(S) - \sum_{i\in S} m_i^c \qquad S<I.$$

考虑到 I 中所有可能的团伙 S 对分摊方案的态度，取

$$v_i = \min_{\substack{S\subset I\\ i\in S}} g^C(S), \quad i=1,2,\cdots,n$$

令权系数为

$$w_i = \frac{v_i}{\sum_{i=1}^{n} v_i}, \quad i=1,2,\cdots,n$$

从而各部门分摊额为

$$M_i^c = m_i^c + w_i g^c(I), \quad i=1,2,\cdots,n$$

3. 非合作对策模型

在非合作对策中，每个部门都希望自己得到的利益最大，因此在考虑费用分摊时，需要通过各部门相互协商或谈判来解决. 在谈判过程中，若各部门能遵守一定的“合理性”假设，那么 Nash 谈判模型的解即为满足这些“合理性”假设的解. 具体地说，假设 u_i 为部门 i 的效用函数，谈判的起点为 $\boldsymbol{d}=(d_1,d_2,\cdots,d_n)$，亦称为现状点，表示谈判破裂时的冲突点，在费用分摊问题中，对应着各部门所愿意(或能够)承担的分配比例的上界. 当 u_i 和 d_i 给定后，Nash 谈判模型的解，即为下面非线性规划问题的最优解：

$$\max \prod_{i=1}^{n} (u_i - d_i)$$
$$\text{s.t.} \quad u \in U$$

对策论模型更多地从相互竞争的角度去考虑问题，在公平合理性方面比传统方法要好得多，因此特别适合于竞争条件下的各部门合作投资的场合. 不足之处在于信息量要求较多，计算较复杂，也缺乏协商.

三、团体决策方法

在许多决策问题中，参与决策的单位（决策者）有时不止一个，这些单位经常是通过他们的代表组成各种委员会或其他形式的决策机构，使每一项决策能尽量满足群体中各单位的要求和愿望. 如何集中团体中各成员的意见构成整个团体的意见，就是团体决策要研究的问题.

首先要做的工作就是研究团体决策的过程及实质，用数学语言给出一个定义，然后以一组公理来描述其合理性，进而寻找“合理”的解答.

考察评选优秀运动员、选举代表和评定啤酒质量等活动的决策过程，这些过程的共同点是，参加评选的每个人（称为选民）对评选对象（称为候选人）有一个排序，而团体决策就是要根据每个选民的排序，根据一种选举规则，确定选举结果. 将团体决策的选举规则定义如下：

用 $I=(1,2,\cdots,n)$ 表示选民集合，用 $A=(x,y,z,\cdots)$ 表示候选人集合. 选举要求每个选民 $i\in I$ 对全体候选人 A 作一排序，记为 P_i. 所谓选举规则是根据每个 $P_i(i=1,2,\cdots,n)$ 确定群体对 A 的排序，记作 P，这种由 $(P_1,P_2,\cdots,P_n)$ 到 P 的对应关系在团体决策中称为团体一致函数.

显然对于任何一个排序 P_i 和 P，必须具备以下两个性质（也称为公理）：

1° 唯一性. $\forall\ x,y\in A$，下面三种关系必有且仅有一种成立：$x>y$（x 优于 y），$x\sim y$（x 等同 y），$x<y$（x 劣于 y）. 用 $\geqslant$ 记优于或等同，$\leqslant$ 记劣于或等同.

2° 传递性. $x,y,z\in A$, 若 $x\geqslant y, y\geqslant z$, 则 $x\geqslant z$.

Arrow 认为要从理论上讨论选举规则的合理性，必须对团体一致函数的性质加上若干限制.

1° A 中元素的个数等于或大于 3 个.

2° 团体一致函数对 $(P_1,P_2,\cdots,P_n)$ 中所有投票都有意义.

3° 选民至少有 2 个.

这些要求显然是合理的.

Arrow 提出"合理"的选举规则应满足四条公理：

公理 1(选民权利公理)　对任一对候选人 x 和 y, 都存在一次投票，根据选举程序能确定 $x>y$.

公理 2(社会评价和个人评价之间正相关)　若对于第一次投票，选举程序确定 $x>y$, 而在第二次每个选民 $\{i\}$ 的投票 P_i 中，x 的次序或者与第一次相同，或者提前，其他候选人的次序不变，那么对于第二次投票，选举程序也应确定 $x>y$.

公理 3(不相关方案的独立性)　设 A_1 是 A 的子集，若在两次投票中，每个选民对 A_1 中各候选人的排序不变，那么在选举程序所确定的两次选举结果中，A_1 中各候选人的排序相同.

公理 4(非独裁)不存在这样的选民 $\{i\}$, 使得对任两个候选人 x 和 y, 只要 $(x>y)_i$, 选举程序就有 $x>y$.

然而 Arrow 证明，符合这种"合理性"的选举规则是不存在的.

在实用中，人们常常引进效用函数的概念，将决策中的个人偏好量化，研究团体效用函数 $u=u(u_1,u_2,\cdots,u_n)$ (u_i 为团体中成员 i 的效用函数，$i=1,2,\cdots,n$) 具有何种可能的形式，以及具有这种形式的条件，从而能够在一定条件下构造一些具体的团体效用函数.

在分摊问题中，一般情况下，各部门在研究分摊比例时，通常都是从本部门利益出发，希望分摊方案尽量对自己有利. 具体说，在二次性分摊法的计算公式中，

$$M_i = m_i + w_i[C(I) - \sum_{i=1}^{n} m_i], \quad i=1,2,\cdots,n$$

其中，M_i 和 m_i 为部门 i 的分摊额和可分费用额，$C(I)$ 为总费用. 剩余费用分摊比例系数 w_i 既要使部门 i 感到合理，可以接受，又要能使整个群体感到满意，从而充分发挥各部门的积极性，产生联合应具有的经济或社会效益.

设 u_i 为部门 i 的效用函数，且设 u_i 只和该部门的剩余费用分摊比例系数 w_i 有关，即

$$u_i = u_i(w_i), \quad i=1,2,\cdots,n,$$

则团体效用函数为

$$u = u(u_1(w_1), u_2(w_2), \cdots, u_n(w_n)),$$

由此可以得到费用分摊问题中，决定剩余费用分摊比例系数 w_i，$i=1,2,\cdots,n$ 的团体决策模型(CAGD)：

$$\max u(u_1(w_1), u_2(w_2), \cdots, u_n(w_n))$$

$$\text{s.t.} \quad 0 \leqslant w_i \leqslant \frac{c_i - m_i}{C(I) - \sum_{i=1}^{n} m_i} \quad i=1,2,\cdots,n$$

$$\sum_{i=1}^{n} w_i = 1$$

约束条件是由分摊额 M_i 不应超过单独投资这一假设而得，即由

$$M_i = m_i + W_i(C(I) - \sum_{i=1}^{n} m_i) \leqslant C_i$$

有

$$w_i \leqslant \frac{c_i - m_i}{C(I) - \sum_{i=1}^{n} m_i} \quad i=1,2,\cdots,n$$

根据效用理论和团体决策理论，可以认为 u 为一向量值函数，即 $u=u(u_1,u_2,\cdots,u_n)$，也可假设效用函数 u 具有可加性，即 $u(w) = \sum_{i=1}^{n} \lambda_i u_i(w_i)$，其中 $\lambda_i \geqslant 0$，$\sum_{i=1}^{n} \lambda_i = 1$. 显然权系数 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 的选

取与问题的解关系密切.因此首先要确定一组实际上表示各部门意见相对重要程度的权系数 $\lambda_i, i=1,2,\cdots,n$,才能使求得的解比较好地集中各部门的意见,让团体和各部门都感到满意.

这里研究利用委托过程确定各部门意见相对重要程度的权系数.

假设团体效用函数 $u(w)=\sum_{i=1}^{n}\lambda_i u_i(w_i)$,且部门 i 的效用函数 u_i 已知,并假设各部门对确定剩余费用分摊比例系数都负有责任,而且可以提出自己的看法,则所采用的分摊方法体现出公平和调和,既集中了各方的意见,又有利于消除分歧.构造的委托过程包含下述三条公理:

公理 1(委托) 团体中每一个部门有一委托小组,这个小组由其余 $n-1$ 个部门组成,部门 i 对其委托小组内的每一个部门 j 指定一个权 P_{ij}, $0\leqslant P_{ij}\leqslant 1, j=1,2,\cdots,n$; $\sum_{j=1}^{n}P_{ij}=1, P_{ii}=0, i=1,\cdots,n$.

权系数 P_{ij} 可以理解为部门 i 认为部门 j 在费用分摊中应承担的责任的多少,也可以理解为部门 j 在研究部门 i 的剩余费用分摊比例时发表意见的客观性和公平性. $P_{ii}=0, i=1,2,\cdots,n$,表明部门 i 在研究自己的分摊比例时,应采用回避政策.

公理 2(决策) 每个委托小组都有一个形如 $u'_i(w)=\sum_{j=1}^{n}P_{ij}u(w_j)$ 的效用函数.

公理 3(代替) 用部门 i 的委托小组的效用函数去代替部门 i 的效用函数 $u_i, i=1,2,\cdots,n$.

委托过程的实施有两种类型.

第一种类型 Pareto 委托过程

设部门 i 的效用函数为 $u_i^0, i=1,2,\cdots,n$. 在前面三公理下,第一步委托过程将得到部门 i 的新效用函数为

$$u_i^1=\sum_{j=1}^{n}P_{ij}u_j^0 \quad i=1,2,\cdots,n$$

一般地,根据向量值效用函数 $u'=(u'_1,u'_2,\cdots,u'_n)$ 得出的结果并不一定令人满意,于是可进行第二步委托,得到

$$u_i^2=\sum_{j=1}^{n}P_{ij}u_j^1, \quad i=1,2,\cdots,n$$

上述委托过程可用矩阵表示. 令 $u^k=(u_1^k,u_2^k,\cdots,u_n^k)^T$,表示在第 k 步委托过程得到的向量值效用函数,令 $P=(P_{ij})_{n\times n}$,则有

$$u^1=Pu^0,\cdots,u^k=Pu^{k-1}=\cdots=P^ku^0$$

可以证明,若存在某个 i,使矩阵 P 的第 i 行 $P_i=(P_{i1},P_{i2},\cdots,P_{in})>0$,则对向量值优化问题:

$$\max u(w)=u(u_1(w),\cdots,u_n(w))$$
$$\text{s.t. } w\in X \quad i=1,2,\cdots,n$$

的解集 X,有 $X_k\subset X_{k-1},k=1,2,\cdots$. 这里 X_k 是关于向量值效用函数 u^k 的 Pareto 解集. 由矩阵 P 的秩$\leqslant n-1$,可知每一步委托得到的 Pareto 解集将逐渐缩小. 当 X_k 缩减到仅剩一个点或较少的点,从而可以选择出最满意方案时,委托过程终止.

第二种类型　Marcov 委托过程

将矩阵 P 看作是具有 n 个状态的 Marcov 过程的一步转移概率矩阵,根据 Marcov 过程的理论,在一定条件下可证明这是马氏链中的正则链,因此一般来讲存在极限分布(或稳定概率分布向量),$\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\cdots,\lambda_n)^T$,使 $\lim\limits_{k\to\infty}P_i^k=\boldsymbol{\lambda}^T$. 此时有

$$\lim_{k\to\infty}u_i^k=u, \quad i=1,2,\cdots,n.$$

其中,u 即为形如 $u=\sum_{j=1}^{n}\lambda_ju_j(w_j)$ 的团体效用函数,相应的权系数 $(\lambda_1,\lambda_2,\cdots,\lambda_n)$ 即为上述极限值 $\boldsymbol{\lambda}^T$,它可以由求解如下非负线性方程组得到

$$\lambda_j=\sum_{i=1}^{n}\lambda_iP_{ij} \quad i=1,2,\cdots,n$$

$$\sum_{j=1}^{n}\lambda_j=1$$

这样就可得到一个反映和综合了各部门意见的效用函数 u，根据 u 值的大小可以对所有方案进行排序.

下面是一个数值例子

例 4-2 三部门费用分摊.

考虑一个有三个部门的费用分摊问题，已知各种可能的合作条件下的投资费用如下（单位：百万元）：

$C(1)=80, C(2)=C(3)=70,$

$C(1,2)=85, C(1,3)=90, C(2,3)=105,$

$C(1,2,3)=110.$

由可分费用的计算公式

$$M_i=C(I)-C(I\setminus\{i\}),\quad i=1,2,\cdots,n$$

知
$$m_1=5, m_2=20, m_3=25.$$

所以剩余费用分摊问题的团体决策模型为

$$\begin{aligned}&\max\ u(u_1(w_1),u_2(w_2),u_3(w_3))\\&\text{s.t.}\ 0\leqslant w_1\leqslant 1, 0\leqslant w_2\leqslant 0.83\\&0\leqslant w_3\leqslant 0.75, w_1+w_2+w_3=1\end{aligned}$$

在第一个约束的计算中，由 $W_i\leqslant\dfrac{C(i)-m_i}{C(I)-\sum_{i=1}^{n}m_i}$ 知，$w_i\leqslant\dfrac{80-5}{110-(5+20+25)}=\dfrac{75}{60}=1.25$，又 $0\leqslant w_i\leqslant 1$，故取 w_1 的上界为 $\min\left\{1,\dfrac{C(I)-m_1}{C(I)-\sum_{i=1}^{n}m_i}\right\}=1.$

在委托过程中，设委托矩阵为

$$P=\begin{bmatrix}0 & 0.4 & 0.6\\0.3 & 0 & 0.7\\0.4 & 0.6 & 0\end{bmatrix}.$$

在 Pareto 委托过程中，效用函数取为

$$u_i(w_i)=1-w_i^2, i=1,2,3$$

在 Marcov 委托过程中，效用函数取为

$$u_i(w_i)=a-be^{-cw_i}, i=1,2,3$$

在确定系数a、b、c时，假定每个部门均属回避风险型的，即取$u_i(0)=1, u_i(1)=0, u_i(0.5)=0.75$，可得 $a=1.125, b=0.125, c=-2.197$，从而

$$u_i(w_i)=1.125-0.125e^{2.197w_i} \quad i=1,2,3.$$

在利用 Nash 谈判模型求解时，u_i 仍取上述形式，现状点是根据各部门所承担的剩余费用分摊比例系数 w_i 的上界给出的，即

$$d_1=d_1(1)=1.125-0.125e^{2.197\times 1}=0$$

$$d_2=d_2(0.83)=1.125-0.125e^{2.197\times 0.83}=0.3508$$

$$d_3=d_3(0.75)=1.125-0.125e^{2.197\times 0.75}=0.4756$$

1° 利用 SCRB(二次性分摊)和 CGA(最小费用缺口法)计算过程简单，略去.

2° 利用 Nash 谈判模型，用非线性规划方法解得 $w_1=0.472$, $w_2=0.305$, $w_3=0.222$，故三个部门的分摊费用为

$$M_1^c=m_1+w_1\left(C(I)-\sum_{i=1}^{3}m_i\right)=5+0.472(110-50)=33.32$$

$M_2^c=38.3$， $M_3^c=38.32$.

3° 利用 Shapleg 方法的计算过程如表 4-1，表 4-2，表 4-3 所示，其中，

$$v(\varphi)=0 \qquad v(\{1\})=v(\{2\})=v(\{3\})=0,$$

$$v(\{1,2\})=C(1)+C(2)-C(1,2)=80+70-85=65,$$

$$v(\{1,3\})=C(1)+C(3)-C(1,3)=80-70-90=60,$$

$$v(\{2,3\})=C(2)+C(3)-C(2,3)=70-70-105=35,$$

$$v(\{1,2,3\})=C(1)+C(2)+C(3)-C(1,2,3)$$

$$=80+70+70-110=110.$$

表 4-1

φ_1 S	{1}	{1,2}	{1,3}	{1,2,3}
$v(S)$	0	65	60	110
$v(S\backslash\{1\})$	0	0	0	35
$\lvert S\rvert$	1	2	2	3
$W(\lvert S\rvert)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
$W(\lvert S\rvert)[v(S)-v(S\backslash\{1\})]$	0	1.083	10	25

$$\therefore \qquad \varphi_1(v)=10.83+10+25=45.83$$

分担费用 $\quad M_1^c=C(1)-\varphi_1(v)=80-45.83=34.17$

剩余费用的分摊比例系数

$$W_1=\frac{M_1^c-m_1}{C(I)-\sum_{i=1}^{1}m_i}=\frac{34.17-5}{110-50}=0.486$$

表 4-2

φ_2 S	{2}	{1,2}	{2,3}	{1,2,3}
$v(S)$	0	65	35	110
$v(S\backslash\{2\})$	0	0	0	60
$v(S)-v(S\backslash\{2\})$	0	65	35	50
$\lvert S\rvert$	1	2	2	3
$W(\lvert S\rvert)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
$W(\lvert S\rvert)[v(S)-v(S\backslash\{2\})]$	0	10.83	5.83	16.67

$$\therefore \qquad \varphi_2(v)=10.83+5.83+16.67=33.3$$

分担费用 $\quad M_2^c=C(2)-\varphi_2(v)=70-33.3=36.7$

剩余费用的分摊比例系数

$$W_2=\frac{M_2^c-m_2}{C(I)-\sum_{i=1}^{2}m_i}=\frac{36.7-20}{110-50}=0.278$$

表 4-3

φ_3 S	{3}	{1,3}	{2,3}	{1,2,3}
$v(S)$	0	60	35	110
$v(S\backslash\{3\})$	0	0	0	65
$v(S)-v(S\backslash\{3\})$	0	60	35	45
$\|S\|$	1	2	2	3
$W(\|S\|)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
$W(\|S\|)[v(S)-v(S\backslash\{3\})]$	0	10	5.83	15

$\therefore$ $\varphi_3(v)=10+5.83+15=30.83$

分担费用 $M_3^c=C(3)-\varphi_3(v)=70-30.83=39.17$

剩余费用的分摊比例系数

$$W_3=\frac{M_3^c-m_3}{C(I)-\sum_{i=1}^{3}m_i}=\frac{39.17-25}{110-50}=0.236.$$

4° Marcov 委托过程

根据线性代数方程组

$$\begin{cases}\lambda_1=0.3\lambda_2+0.4\lambda_3,\\ \lambda_2=0.4\lambda_1+0.6\lambda_3,\\ \lambda_3=0.6\lambda_1+0.7\lambda_2,\\ \lambda_1+\lambda_2+\lambda_3=1.\end{cases}$$

解得 $\lambda_1=0.262,\lambda_2=0.342,\lambda_3=0.396.$

所以团体效用函数为

$$u=0.262(1.125-0.125e^{2.197W_1})+0.342(1.125-0.125e^{2.197W_2})+0.396(1.125-0.125e^{2.197W_3})$$

求解非线性规划问题

$$\max u(W)$$

$$\text{s.t. } W_1+W_2+W_3=1$$

得 $$W_1=0.438, W_2=0.315, W_3=0.248$$

三个部门的分摊费用分别为

$$M_1^c=m_1+W_1\left(C(I)-\sum_{i=1}^{3}m_i\right)=5+0.438(110-50)=31.28$$

$$M_2^c=m_2+W_2\left(C(I)-\sum_{i=1}^{3}m_i\right)=20+0.315(110-50)=38.9$$

$$M_3^c=m_3+W_3\left(C(I)-\sum_{i=1}^{3}m_i\right)=25+0.248(110-50)=39.88$$

将上述不同方法得到的费用分摊结果列于表 4-4.

表 4-4

方法	剩余费用分摊比例系数			分摊额			总额
	W_1	W_2	W_3	M_1	M_2	M_3	
SCRB	0.446	0.294	0.262	31.47	37.65	40.85	110
CGA	0.387	0.323	0.29	28.23	39.35	42.42	110
Shapley 值	0.486	0.278	0.236	34.16	36.67	39.17	110
Nash 谈判法	0.472	0.305	0.222	33.32	38.3	38.32	110
Pare 后委托	0.413	0.315	0.272	29.78	38.9	41.32	110
Marcov 委托	0.438	0.315	0.248	31.28	39.9	39.88	110
可分费用				5	20	25	50

从表中可以看出，传统方法、对策方法和团体决策方法的结果在趋势上是一致的，而各种方法也体现出不同的特点.

习　题

1. 甲拥有一幢私房，若该房由甲自己使用，每年可获利 1000 元，若转让给乙使用，每年可获利 2000 元，若转让给丙使用，每年可获利 2500 元，若三人共同使用，则可获利 3000 元，显然以三人共同使用获利最多. 试计算一下三人应如何分配所得的 3000 元.

2. 会议有 100 个席位，分别为三个党派所得. 它们分别拥有 34，33，33 个席位，设法律规定提案需过半数才算通过，现有一提案全票通过，试计算一下各党派在这次提案通过中所起的作用.

3. 理事会有五个常任理事和十个非常任理事，提案仅当全部常任理事和至少四个非常任理事赞成时才通过，求每位常任理事和每位非常任理事在投票中的权重.

4. 奇数个席位的理事会由三派组成，议案表决实行半数通过方案，证明在任一派都不能操纵表决的条件下，三派占有的席位不论多少，他们在表决中的权重是一样的.

5. 100 人的团体由四派组成，人数分别为 40、30、20、10 人. 决议需过半数成员赞成方可通过. 每个派别成员同时投赞成或反对票，用 Shapley 值方法计算各派在表决中的权重.

第五章　常微分方程

在实际问题中，常常需要知道两个变量之间的关系，或一个变量随另一个变量变化的规律，常微分方程是利用机理分析方法研究这一问题的重要工具．其主要特点是利用微元分析法，建立瞬时变化率的表达式，然后根据所给条件，确定解曲线．因此，对“变化率”的假设与推导，是建立常微分方程模型的关键．

§5.1　常微分方程模型的建模步骤

我们以一个例子来说明常微分方程模型建立的基本步骤．

例 5-1　某人的食量是 10467 焦/天，其中 5038 焦/天，用于基本的新陈代谢(即自动消耗)．在健身训练中，他所消耗的热量大约是 69 焦/公斤·天乘以他的体重(公斤)．假设以脂肪形式贮藏的热量 100%地有效，而 1 公斤脂肪含热量 41868 焦．试研究此人的体重随时间变化的规律．

分析　题中并未出现“变化率”、“导数”这样的关键词，但要寻找的是体重(记为 W)关于时间 t 的函数．如果把体重 W 看作是时间 t 的连续可微函数，我们就能找到一个含有$\frac{\mathrm{d}W}{\mathrm{d}t}$的微分方程．

问题中涉及的时间仅仅是“每天”，由此，对于“每天”．

体重的变化＝输入－输出

其中输入是指扣除了基本新陈代谢之后的净重量吸收；输出是进行健身训练时的消耗(WPE)．

由于考虑的是导数，因此，上述陈述可以表示为更好的结构式

体重的变化/天＝净吸收量/天－WPE/天

净吸收量/天＝10467(焦/天)－5038(焦/天)

$$=5429(焦/天)$$

$$净输出量/天=69(焦/公斤\cdot天)\times W(公斤)$$

$$=69W(焦/天)$$

$$体重的变化/天=\frac{\Delta W}{\Delta t}(公斤/天)\underset{\Delta t\to 0}{=}\frac{\mathrm{d}W}{\mathrm{d}t}$$

这就是所需要的关于连续函数 $W(t)$的瞬时关系.注意到有些量是用能量(焦)的形式给出的,而另外一些量是用重量的形式(公斤)给出,考虑单位的匹配,利用

$$公斤/天=\frac{净焦/天}{41868\ 焦/公斤}$$

因此有

$$\frac{\mathrm{d}W}{\mathrm{d}t}=\frac{(2500-1200)-16W}{10000}$$

这个式子可用物理单位(量纲)检查如下:

$$\frac{公斤}{天}=\frac{焦/天-(焦/公斤\cdot天)\cdot公斤}{焦/公斤}$$

上述微分方程是一阶线性的,答案中有一个积分常数,因此有一个定解条件.例如,一天开始时他的体重为

$$W|_{t=0}=W_0$$

至此,已建立起问题的常微分方程模型.为了解模型所说明的问题,用分离变量法求解

$$\frac{\mathrm{d}W}{1300-16W}=\frac{\mathrm{d}t}{10000}$$

$$-\frac{1}{16}\ln|1300-16W|=\frac{t}{10000}+c$$

利用所给初始条件

$$c=-\frac{1}{16}\ln|1300-16W_0|$$

从而得到

$$|1300-16W|=|1300-16W_0|\exp(-16t/10000)$$

因为指数因子为正,所以$(1300-16W)$与$(1300-16W_0)$同

号，故可将绝对值符号去掉，因此有

$$1300-16W=(1300-16W_0)\exp(-16t/10000)$$

解出 W，

$$W=\frac{1300}{16}-\left(\frac{1300-16W_0}{16}\right)\exp(-16t/10000)$$

显然，当 $t\to\infty$ 时，体重有稳定值

$$W_{平衡}=\frac{1300}{16}=81.25$$

至此，问题已基本上解决.

回顾上述过程，可以将常微分方程模型的建立步骤整理如下：

1° 翻译或转化. 在实际问题中，有许多表示导数的常用词，如"速率"、"增长"(在生物学以及人口问题研究中)，"衰变"(在放射性问题中)，以及"边际的"(在经济学中)等.

2° 建立瞬时表达式. 根据自变量有微小改变 Δt 时，因变量的增量 ΔW，建立起在 Δt 时段上的增量表达式，令 $\Delta t\to 0$，即得到 $\frac{dW}{dt}$ 的表达式.

3° 配备物理单位. 在建模中应注意每一项采用同样的物理单位.

4° 确定条件. 这些条件是关于系统在某一特定时刻或边界上的信息，它们独立于微分方程而成立，用以确定有关的常数. 为了完整充分地给出问题的数学陈述，应将这些给定的条件和微分方程一起给出.

上面是利用微元法建立微分方程模型的基本步骤. 建立常微分方程模型还有其它方法，下面是常用的两种.

1° 按变化规律直接列方程，即利用人们熟悉的力学、数学、物理、化学等学科中的规律，如牛顿第二定律，放射性物质的放射规律等，对某些实际问题直接列出微分方程.

2° 模拟近似法. 在生物、经济等学科中，许多现象所满足的规律并不很清楚，而且现象也相当复杂，因而需要根据实际资料或大

量的实验数据，提出各种假设. 在一定的假设下，给出实际现象所满足的规律，然后利用适当的数学方法得出微分方程. 这一方法的实例可参考第七章水塔流量问题中水流速率的曲线拟合和常微分方程反问题的参数估计.

§5.2 肿瘤的生长规律

利用数学模型研究恶性肿瘤的生长规律，是人类对癌症研究的一个方面，它有助于人类认识其生长规律，寻找控制消灭它的措施.

通过临床观察，人们发现肿瘤细胞的生长有下列现象：

1° 按照现有手段，肿瘤细胞数目超过 10^{11}时，临床才可能观察到.

2° 在肿瘤生长初期，每经过一定的时间，肿瘤细胞数目就增加一倍.

3° 在肿瘤生长后期，由于各种生理条件的限制，肿瘤细胞数目逐渐趋向某个稳定值.

一、指数模型

假设：肿瘤细胞的增长速度与当时该细胞数目成正比，比例系数(相对增长率)为 λ. 设时刻 t 肿瘤细胞数目为 $n(t)$，则

$$\frac{\mathrm{d}n}{\mathrm{d}t}=\lambda n$$

解为

$$n(t)=c\mathrm{e}^{\lambda t}$$

据临床观察 1°，可令 $n(0)=10^{11}$，得肿瘤细胞生长规律为

$$n(t)=n(0)\mathrm{e}^{\lambda t}=10^{11}\mathrm{e}^{\lambda t}$$

据临床观察 2°，设细胞增加一倍所需时间为 τ，则，

$$n(t+\tau)=2n(t)$$

从而得

$$\tau=\ln 2/\lambda$$

该模型未能反映出临床观察 3°,在人口问题中,称这种指数模型为马尔隆斯(Malthus)模型.

二、Verhulst 模型

考虑到临床观察 3°,对指数模型进行修正. 由此,荷兰生物数学家 Verhulst 提出设想:相对增长率随细胞数目 $n(t)$ 的增加而减少. 若用 N 表示因生理限制肿瘤细胞数目的极限值, $f(n)$ 表示相对增长率,则 $f(n)$ 为 $n(t)$ 的减函数,为处理方便,令 $f(n)$ 为 n 的线性函数:

$$f(n)=a+bn.$$

假设当 $n(t)=N$ 时, $f(n)=0$, 当 $n(t)=n(0)$ 时, $f(n)=\lambda$, 可得相对增长率为

$$f(n)=\lambda\cdot\frac{N-n(t)}{N-n(0)}$$

则 $n(t)$ 满足微分方程

$$\frac{\mathrm{d}n}{\mathrm{d}t}=\lambda n(t)\frac{N-n(t)}{N-n(0)}$$

解为

$$n(t)=n(0)\left[\frac{n(0)}{N}+\left(1-\frac{n(0)}{N}\right)\mathrm{e}^{-\frac{\lambda Nt}{N-n(0)}}\right]^{-1}.$$

在实际应用中,常用的是上述方程的一个特殊形式.

$$f(n)=\lambda\frac{N-n(t)}{N},$$

相应微分方程为

$$\frac{\mathrm{d}n}{\mathrm{d}t}=\lambda\left(1-\frac{n(t)}{N}\right)n(t).$$

该方程称为 Logistic 模型或者 Verhulst-Pearl 阻滞方程,广泛应用于医学、农业、生态和商业等领域. 为此再介绍另外两种推导方

法.

1°令 λ 是大小为 n 的肿瘤的潜在增长率(Potential rate),即如果没有生理限制而且细胞之间互不影响,那么细胞数目就按这个潜在率增长,这里 λ 是自然增长的固有率.现在假设实际的增长率是这个潜在率乘上一个比例——最大可能的细胞数目 N 中还未出生部分所占的比例,即 $\frac{N-n(t)}{N}$,可得微分方程.

$$\frac{\mathrm{d}n}{\mathrm{d}t}=\lambda\frac{N-n(t)}{N}\cdot n(t).$$

2°Lotka(1925)提出的推导如下:设相对增长率 $f(n)$ 能展开成 $n(t)$ 的幂级数,则

$$\frac{\mathrm{d}n}{\mathrm{d}t}=c_0+c_1n+c_2n^2+\cdots$$

在肿瘤问题中(或在生态问题中),显然 $n=0$ 时,$\frac{\mathrm{d}n}{\mathrm{d}t}=0$,于是 $c_0=0$.

若令 $\frac{\mathrm{d}n}{\mathrm{d}t}=c_1n$,$c_1=\lambda$,即为指数模型.仅在 $n=0$ 时,$\frac{\mathrm{d}n}{\mathrm{d}t}=0$,而对所有其他的细胞数目,都有 $\frac{\mathrm{d}n}{\mathrm{d}t}>0$.现在要求有两个零点,即当 $n(t)=N$ 时,也应有 $\frac{\mathrm{d}n}{\mathrm{d}t}=0$,满足这个要求的最简单形式是级数终止于 n^2 项,即

$$F(n)=\frac{\mathrm{d}n}{\mathrm{d}t}=c_1n+c_2n^2.$$

$F(n)$ 的根为 $n_1=0$ 和 $n_2=-\frac{c_1}{c_2}$,由 $c_1=\lambda$ 和 $n_2=-\frac{c_1}{c_2}=N$,有 $c_1+c_2N=0$,$c_2=-\frac{\lambda}{N}$,所以微分方程为

$$\frac{\mathrm{d}n}{\mathrm{d}t}=\lambda\left(1-\frac{n}{N}\right)n.$$

利用分离变量法求解上述微分方程,

$$\frac{\mathrm{d}n}{n\left(1-\frac{n}{N}\right)}=\lambda\mathrm{d}t$$

$$\therefore \quad \ln n-\ln(N-n)=\lambda t+c \quad \frac{n}{N-n}=ce^{\lambda t}$$

$$\text{而} \quad n(0)=n_0, \quad \therefore \frac{n_0}{N-n_0}=c$$

$$\therefore \quad n(t)=n_0e^{\lambda t}\left[1+\frac{n_0}{N}(e^{\lambda t}-1)\right]^{-1}.$$

三、Gompertzlan 模型

由于 Verhulst 模型与某些实测数据吻合不好，考虑将相对增长率从 $n(t)$ 的线性函数修改为 $n(t)$ 的对数函数，即相对增长率为

$$-\lambda\ln\frac{n}{N}$$

其中，负号表示随 $n(t)$ 的增加而减少，但不是线性关系，而是与 $n(t)$ 在极限值中所占比例的对数有关. 得微分方程

$$\frac{dn}{dt}=-\lambda n\ln\frac{n}{N},$$

其解为

$$n(t)=n(0)\exp\left[\left(\ln\frac{N}{n(0)}\right)(1-e^{-\lambda t})\right]=n(0)\left[\frac{N}{n_{(0)}}\right]^{1-e^{-\lambda t}}.$$

为了将这三个模型进行比较，在 t-$\ln n(t)$ 半对数坐标中画出了它们的曲线. 从中可以看到 Gompertzlan 模型的曲线增长较快.

半世纪 80 年代，有人对肿瘤生长规律提出了更一般的模型：

$$\frac{dn}{dt}=\lambda_n\frac{1-\left(\frac{n}{N}\right)^{\alpha}}{\alpha},\alpha\geqslant 0,$$

其解为

$$n(t)=n(0)\left\{\left(\frac{n(0)}{N}\right)^{\alpha}+e^{-\lambda t}\left[1-\left(\frac{n(0)}{N}\right)^{\alpha}\right]\right\}^{-\frac{1}{2}}$$

显然，当 $\alpha\to 1$ 且 $N\to\infty$ 时，$\frac{dn}{dt}\to\lambda n$，即为指数模型. 当 $\alpha\to 1$，N 为定值时，$\frac{dn}{dt}\to\lambda n\left(1-\frac{n}{N}\right)$，即为 Logistic 模型.

对上述三种模型，利用试验数据作参数估计参见第七章§7.2.

§5.3 传染病模型

传染病的流行，至今仍威胁着人类.然而，人们在研究传染病的蔓延过程时，却遇到不少困难.这些困难主要是：第一，对传染病的实验极其昂贵(用动物作试验)，而且从道德角度不允许用人作试验；第二，关于传染病的有关数据只能取自于疾病爆发后的有关报告，而报告中的数据往往不全面，并且要准确估计有关参数是很困难的，通常人们仅仅能获得参数的变化范围.因此，数学模型和计算机模拟成为人们研究传染病蔓延过程的重要手段(并不是探讨医学上的传染机理).这个问题和自然科学中一些已经有确定规律的问题不同，不可能立即对它做出恰当的假设，建立完善的模型，只能先做出最简单的假设，建立模型，得出结果，分析是否符合实际，然后针对其不合理或不完善处，进行修改或补充假设，逐步得到较为合理的模型.

模型 1 假设病人通过空气、食物等将病菌传播给健康人.单位时间内一个病人能传染的人数是常数 k_0. 设 t 时刻的病人数目为 $i(t)$，是 t 的连续可微函数，则由假设知

$$i(t+\Delta t)-i(t)=k_0 i(t)\Delta t$$

或

$$\frac{\mathrm{d}i}{\mathrm{d}t}=k_0 i(t)$$

设开始观察时有 i_0 个病人，即 $i(t)|_{t=0}=i_0$，则方程满足初始条件的解为

$$i(t)=i_0 e^{k_0 t}$$

由此看来，病人数目随着时间的推移将无限增加，这与实际情况不符.因为在不考虑疾病流行期间的出生、死亡和迁移时，一个

地区的总人数大致可认为是常数，而 k_0 是变化的，在传染病流行的初期，k_0 较大，随着病人的增多，健康人减少，被传染的机会也将减少，k_0 逐渐变小. 所以对原假设要进行修改.

模型 2 将人群分为两类，病人 $i(t)$ 和健康人 $s(t)$. 假设：1°该地区总人数为 n，且

$$i(t)+s(t)=n$$

2°单位时间内，一个病人传染的人数与当时健康者人数成正比，比例系数为 k（称为传染系数），则

$$\frac{\mathrm{d}i}{\mathrm{d}t}=ks(t)\cdot i(t)$$

或

$$\frac{\mathrm{d}i}{\mathrm{d}t}=k(n-i)i \qquad i(0)=i_0$$

用初等积分法易得方程的解为

$$i(t)=n\Big/\left[1+\left(\frac{n}{i_0}-1\right)\mathrm{e}^{-knt}\right]$$

显然，$i(t)$ 单调增，且当 $t\to\infty$ 时，$i(e)\to n$，即最终所有的人都要被传染. 这与实际情况不符. 但在传染病流行的前期这个模型还是可用的，传染病学者曾用它来预报传染病高潮到来的时刻，即病人人数增加最快的时刻，记

$$u(t)\triangleq\frac{\mathrm{d}x}{\mathrm{d}t}=\frac{kn^2\left(\frac{n}{i_0}-1\right)\mathrm{e}^{-knt}}{\left[1-\left(\frac{n}{i_0}-1\right)\mathrm{e}^{-knt}\right]^2}$$

使 $u(t)=\frac{\mathrm{d}x}{\mathrm{d}t}$ 达到最大值的时刻 t_0 即是传染病高潮到来的时刻，由 $\frac{\mathrm{d}n}{\mathrm{d}t}=0$，求得

$$t_0=\left[\ln\left(\frac{n}{i_c}-1\right)\right]/kn$$

其中，传染系数 k 可由统计资料求得，或根据经验估计.

模型 3 将人群分为三类：病人 $i(t)$、健康人 $s(t)$，以及病愈免疫和死亡者 $r(t)$，假设

1° 设总人数为 n，且 $i(t)+s(t)+r(t)=n$.

2° 同模型 I 的假设 2.

3° 在单位时间内，病愈免疫(包括死亡)的人数 $r(t)$ 与当时病人人数成正比，设比例系数为 l，称 l 为恢复系数，即

$$\frac{dr}{dt}=li(\rho)$$

由假设 2，

$$\frac{di}{dt}=ks(t)i(t)-\frac{dr}{dt}$$

这时模型为

$$\begin{cases}\dfrac{di}{dt}=ksi-li\\[2mm] \dfrac{ds}{dt}=-ksi\end{cases}$$

设初始条件为

$$\begin{cases}i|_{t=0}=i_0\\ s|_{t=0}=s_0=n-i_0 \quad (\text{即 } r|_{t=0}=0)\end{cases}$$

方程组的解析解难以求得，仅在相平面上讨论解的性态.

由方程组知

$$\frac{di}{ds}=\frac{ksi-li}{-ksi}=\frac{li}{ksi}-1\triangleq\frac{\rho}{s}-1,$$

其中，$\rho=\dfrac{l}{k}$ 称为特征指数，对于同一地区、同一种传染病，ρ 是常数. 其初始条件为

$$i|_{s=s_0}=n-s_0$$

方程的解为

$$i=\rho\ln\frac{s}{s_0}-s+n$$

在相平面上过点 (s_0,i_0) 的这条相轨线如图 5-1 所示，因为 $i(t)\geqslant 0$，故图中实线部分有意义. 又因 $\dfrac{ds}{dt}\leqslant 0$，故图中箭头方向表示 t 增加

时，$s(t)$和 $i(t)$的变化趋势.

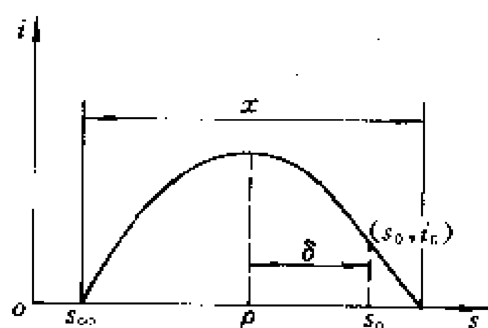


图 5-1

由方程知，当 $s=\rho$ 时，$\frac{di}{dt}=0$，i 达到极大值. 从图上可见，当 $s_0>\rho$ 时，$i(t)\nearrow$；当 $s_0\leqslant\rho$ 时，$i(t)\swarrow$. 这说明，仅当传染病开始时健康人数超过 ρ 的情况下，传染病才会蔓延. ρ 是一个阈值(俗称门槛). 通常 i_0 很小，可近似认为 $s_0\approx n$，在总人数 n 不变的情况下，提高门槛 ρ 的数值，对制止传染病的蔓延有利，这就要使恢复系数 l 增大，传染系数 k 降低，即要提高该地区的医疗水平和卫生保健水平.

可以利用统计数据检验模型 3. 实际上，可以根据医院每周病愈和死亡人数的统计资料，确定模型中的$\frac{dr}{dt}$(t 以周为单位)，同时又可根据上述模型，找出$\frac{dr}{dt}$的近似表达式，将两者进行比较.

由方程$\frac{dr}{dt}=li(t)$和$\frac{ds}{dt}=-ks_i(t)$得

$$\frac{ds}{dr}=-\frac{k}{l}s=-\frac{s}{\rho}.$$

它在初始条件 $s|_{r=0}=s_0$ 下的解为

$$s=s_0e^{-r/\rho}$$

又 $i(t)+s(t)+r(t)=n$，故

$$\frac{dr}{dt}=l(n-r-s_0e^{-r/\rho})$$

当 $r\ll\rho$ 时，利用指数函数的 Taylor 展式可得

$$\frac{dr}{dt}=l[n-s_0+(\frac{s_0}{\rho}-1)r-\frac{s_0}{2\rho^2}r^2]$$

在零初始条件下的解为

$$r(t)=\frac{\rho^2}{s_0}[(\frac{s_0}{\rho}-1)+\alpha\text{th}(\frac{\alpha lt}{2}-\varphi)]$$

其中

$$\begin{cases}\alpha=\sqrt{(\frac{s_0}{\rho}-1)^2+\frac{2s_0(n-s_0)}{\rho^2}}; \\ \varphi=\text{th}^{-1}\frac{s_0/\rho-1}{\alpha},\end{cases}$$

得到

$$\frac{\mathrm{d}r}{\mathrm{d}t}=\frac{\rho^2\alpha^2 l}{2s_0\text{ch}^2(\frac{\alpha lt}{2}-\varphi)}.$$

Kermaqk 和 Mckendrick 利用本世纪初在印度孟买发生的一次瘟疫中的死亡人数检验这个模型，取定参数 ρ, s_0 等的数值后，有

$$\frac{\mathrm{d}r}{\mathrm{d}t}=\frac{890}{\text{ch}^2(0.2t-3.4)}$$

此式表示的曲线如图 5-2 所示，图中的圆点是那次瘟疫中每周的实际死亡人数. 可以看出，理论曲线与实际数据吻合得相当好.

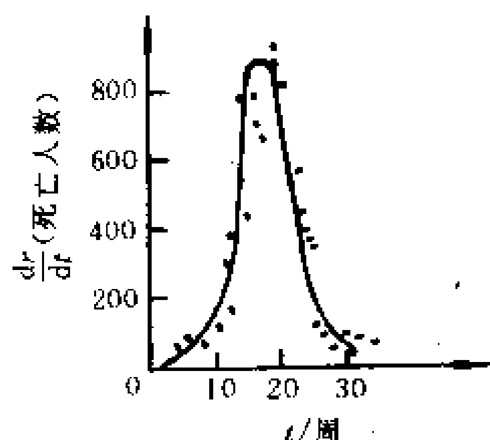


图 5-2　理论曲线与实际数据

其中微分方程的参数估计问题参见第七章.

模型 4　考虑该地区的人口为流动的情况，假设在单位时间内，

1° 有 μ 名健康人进入该地区，流出该地区的人有健康人、病人，他们分别与健康人、病人成正比，比例系数 $\alpha=\mu/n$，n 是总人数；

2° 一个病人能传染的人数与当时的健康人数成正比，比例系数 k 称为传染系数；

3° 健康人因防预而减少发病的人数与健康人数成正比，比例系数 λ 称为预防系数；

4° 病愈免疫的人、死亡的人与病人成正比，比例系数 β 称为恢复系数，d 为死亡系数. 应用人口守恒，有

$$\begin{cases}\dfrac{ds}{dt}=-ksi-(\alpha+\lambda)s+\mu\\ \dfrac{di}{dt}=ksi-\beta i-\alpha i\\ \dfrac{dr}{dt}=-\alpha r+\beta i\end{cases}$$

对上述模型，将第一、二方程联立，可以化为二维问题讨论. 若令 $\rho=(\beta+\alpha)/k$，称为阈值，有下述阈值定理：

定理 1° 当 $\mu/(\alpha+\lambda)\leqslant\rho$ 时，有唯一平衡点 $A(\mu/(\alpha+\lambda),0)\in D$，$D$ 为可行域 $\{(s,i)|s>\beta/k,i\geqslant 0\}$，它是全局扇形稳定的；

2° 当 $\mu/(\alpha+\lambda)>\rho$ 时，有两个平衡点 $A(\mu/(\alpha+\lambda),0)$ 和 $B(\rho,\dfrac{\mu}{\alpha+d}-\dfrac{\alpha+\lambda}{\alpha+d\rho})$，其中 A 是不稳定的平稳点，B 是全局稳定的.

讨论可知，在 ρ 和 μ 不变的假设下，要求传染病不流行，即 $\dfrac{\mu}{\alpha+\lambda}\leqslant\rho$，加强预防隔离措施，增强预防系数 λ 是有效的；当 $\dfrac{\mu}{\alpha+\lambda}>\rho$，且平衡点 B 是焦点，传染病发病有第一高峰期，还可以有第二高峰期.

§5.4 糖尿病的检测

糖尿病是一种新陈代谢疾病，临床表现为血液和尿中含有过多的糖. 它是胰岛素缺少引起的新陈代谢紊乱所致，糖尿病的诊断是根据葡萄糖容许剂量测试(GTT)的结果而作出结论的. 诊断中所遇到的一个困难是对于葡萄糖容许剂量的标准看法不同. 60 年代中期，北爱尔兰 Mayo 医院的 Roseveor 和 Molner 医生以及明尼苏达大学的 Ackerman 和 Gatewood 博士研究了血糖循环系统，建立了数学模型，为糖尿病的诊断提供了比较可靠的根据.

根据生物、医学原理，作如下假设：

1° 葡萄糖在任何有脊椎生物的新陈代谢中起着十分重要的作用，它是所有细胞和组织的能量来源. 每个人都有最适当的血糖浓度，超过这个浓度时，将导致疾病甚至死亡.

2° 血糖浓度在其自我调节过程中，受到生理激素和其他新陈代谢因素的影响和控制.

胰岛素：人吃下碳水化合物后，肠胃系统向胰岛发出信号，使之分泌更多的胰岛素. 此外，血液中的葡萄糖直接刺激胰腺分泌胰岛素. 胰岛素的作用在于促进组织对葡萄糖的吸收.

胰高血糖素：其作用是加快糖原向葡萄糖分解的速度. 低血糖促进胰高血糖素的分泌，而高血糖则抑制它的分泌.

肾上腺素，它是某种紧急机制的一部分，在极端低血糖时，它能很快增加血液中的葡萄糖浓度. 增加糖原向葡萄糖分解的速度，直接抑制肌肉组织对葡萄糖的吸收，抑制胰岛素分泌，还有助于肝脏内乳酸向葡萄糖转化.

糖皮质激素，在碳水化合物的代谢中起着重要的作用.

甲状腺素，有助于肝脏从甘油、乳酸，以及氨基酸这样的非碳水化合物中生成葡萄糖.

生长激素，它不仅直接影响葡萄糖水平，而且还有对抗胰岛素的作用.

Ackerman 等人试图构造一个模型，来描述 GTT 过程中的血糖调节系统. 因此将注意力集中在两个浓度，即血液中的葡萄糖浓度(用 G 表示)和激素的净浓度(用 H 表示)，后者表示所有有关激素的总效果. 这样一个简化模型仍然能够提供一个关于血糖调节系统的准确描述，这有两方面原因. 第一，研究表明，在正常或接近正常的情况下，一种激素，即胰岛素，与血糖的相互作用如此占优势，以至于一个简单的“浓缩参数模型”就足够了. 第二，有证据指出，血糖量正常不必依赖于血糖调节系统的每一个动力学机制都正常，它依赖于整个调节系统的工作情况，这个系统受着胰岛素

——葡萄糖相互作用的控制. 血糖循环规律如图 5-3 所示.

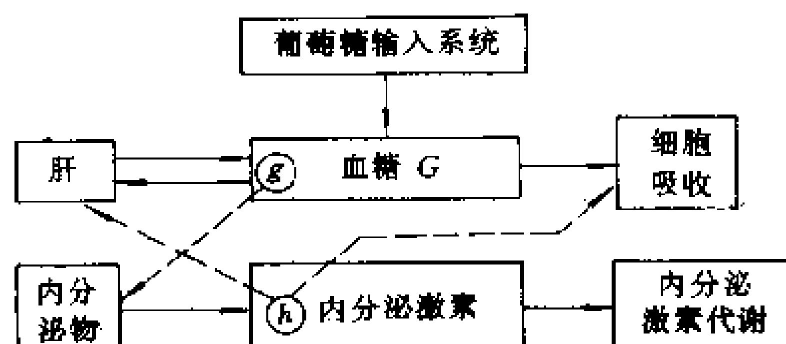


图 5-3 血糖调节系统的简化模型

根据上述假设,血糖浓度的变化依赖于体内原有血糖浓度和内分泌激素浓度,把这种依赖关系记为函数 $F_1(G,H)$,血糖浓度还与外加葡萄糖有关(进行 GTT 测试前患者需服用葡萄糖)、记为 $J(t)$,而内分泌激素浓度的变化则依赖于体内血糖浓度和内分泌激素浓度,记为 $F_2(G,H)$.

这个基本模型由下列方程组给出,

$$\frac{\mathrm{d}G}{\mathrm{d}t}=F_1(G,H)+J(t);$$

$$\frac{\mathrm{d}H}{\mathrm{d}t}=F_2(G,H).$$

假定 $t=0$ 时,$J(t)=0$,患者体内血糖浓度为 G_0,内分泌激素浓度为 H_0,(G_0,H_0 为常量),故有

$$\begin{cases}\left.\dfrac{\mathrm{d}G}{\mathrm{d}t}\right|_{t=0}=0;\\[2ex]\left.\dfrac{\mathrm{d}H}{\mathrm{d}t}\right|_{t=0}=0,\end{cases}\qquad \text{即}\quad \begin{cases}F_1(G_0,H_0)=0;\\F_2(G_0,H_0)=0.\end{cases}$$

我们感兴趣的是 G 和 H 对 G_0,H_0 的偏差,作变换

$$g=G-G_0,\quad h=H-H_0$$

那么

$$\begin{cases}\dfrac{\mathrm{d}g}{\mathrm{d}t}=F_1(G_0+g,H_0+h)+J(t);\\[2ex]\dfrac{\mathrm{d}h}{\mathrm{d}t}=F_2(G_0+g,H_0+h),\end{cases}$$

将 $F_1(G_0+g,H_0+h)$ 和 $F_2(G_0+g,H_0+h)$ 在 (G_0,H_0) 点展开，略去高阶无穷小，则有

$$\begin{cases}\dfrac{\mathrm{d}g}{\mathrm{d}t}=\dfrac{\partial F_1(G_0,H_0)}{\partial H}g+\dfrac{\partial F_1(G_0,H_0)}{\partial H}h+J(t)\\[2ex]\dfrac{\mathrm{d}h}{\mathrm{d}t}=\dfrac{\partial F_2(G_0,H_0)}{\partial G}g+\dfrac{\partial F_2(G_0,H_0)}{\partial H}h\end{cases}$$

我们无法确定四个偏导数的值，但可以确定其符号、由图 5-3 给出的血糖循环过程，可知当 $g>0,h>0$ 时，患者服用葡萄糖，细胞吸收葡萄糖并把它转变为糖元贮藏在肝脏里，血糖浓度降低，所以 $\dfrac{\partial F_1(G_0,H_0)}{\partial G}<0$；又当 $h>0$ 时，细胞吸收葡萄糖并加快葡萄糖转变为糖元的速度，即降低血糖浓度，所以 $\dfrac{\partial F_1(G_0,H_0)}{\partial H}<0$. 综上，$\dfrac{\mathrm{d}g}{\mathrm{d}t}<0$. 在 $g>0$ 时，内分泌器官加速分泌激素，促使 H 增加，有 $\dfrac{\partial F_2(G_0,H_0)}{\partial G}>0$，而通过激素的新陈代谢作用，血液中激素浓度将降低，所以 $\dfrac{\partial F_2(G_0,H_0)}{\partial H}<0$.

将方程组改写为

$$\begin{cases}\dfrac{\mathrm{d}g}{\mathrm{d}t}=-m_1g-m_2h+J(t) & (1)\\[2ex]\dfrac{\mathrm{d}h}{\mathrm{d}t}=-m_3h+m_4g & (2)\end{cases}$$

其中，$m_i(i=1,2,3,4)$ 均为正常数，由于仅需测定血糖浓度，消去 h.

(1)式对 t 求导，得

$$\frac{\mathrm{d}^2g}{\mathrm{d}t^2}=-m_1\frac{\mathrm{d}g}{\mathrm{d}t}-m_2\frac{\mathrm{d}h}{\mathrm{d}t}+\frac{\mathrm{d}J}{\mathrm{d}t}$$

将(2)式代入上式，得

$$\frac{\mathrm{d}^2g}{\mathrm{d}t^2}=-m_1\frac{\mathrm{d}g}{\mathrm{d}t}+m_2m_3h-m_4m_2g+\frac{\mathrm{d}J}{\mathrm{d}t}\qquad(3)$$

又由(1)知 $m_2h=-\dfrac{\mathrm{d}g}{\mathrm{d}t}-m_1g+J(t)$ 代入(3)，得

$$\frac{d^2g}{dt^2}+(m_2+m_3)\frac{dg}{dt}+(m_1m_3+m_2m_4)g=m_3J+\frac{dJ}{dt},$$

改写为

$$\frac{d^2g}{dt^2}+2a\frac{dg}{dt}+w_0^2g=s(t) \tag{4}$$

其中，$a=\frac{1}{2}(m_2+m_3)$，$w_0^2=m_1m_3+m_2m_4$，$s(t)=J(t)+\frac{dJ}{dt}$.

注意到，除了注入葡萄糖的一个极短的时间间隔外，在其他时刻(4)式的右边均为零，利用 Dirac 的 δ-函数，可以有效地研究它. 此外，为了我们的目的，取葡萄糖完全被吸收的时刻 $t=0$，则对于 $t>0$，$g(t)$满足二阶线性齐次方程

$$\frac{d^2g}{dt^2}+2a\frac{dg}{dt}+w_0^2g=0 \tag{5}$$

这个方程具有正系数. 因此，当 $t\to\infty$ 时，$g(t)\to 0$ 即 $G\to G_0$，即模型在预言血糖浓度最终趋于初始浓度这一点上，与实际一致.

$g(t)$的解有三种形式，它取决于 $a^2-w_0^2$ 的符号. 当 $a^2-w_0^2<0$ 时，得

$$g(t)=Ae^{-at}\cos(wt-\delta)$$

其中，$w^2=w_0^2-a^2$，因此

$$G(t)=G_0+Ae^{-at}\cos(wt-\delta)$$

式中有 5 个未知量 G_0，A，a，w_0 和 δ，用下述方法可以确定它们.

在外加葡萄糖被吸收之前，可直接测定患者体内血糖浓度 G_0. 然后，取 $t=t_i(i=1,2,\cdots,n)$，$n\geqslant 4$，利用非线性最小二乘法，选取 A、a、w_0 和 δ，使

$$\min E=\sum_{j=1}^{n}[G_j-G_0-Ae^{-at_j}\cos(wt_j-\delta)]^2$$

在数学实验中，Ackerman 等人发现，测量 G 时，微小的测量误差将导致 a 值产生很大误差，因此，任何含有参数 a 的诊断标准都不可靠，而系统的自然频率 w_0 对测量 G 时的测量误差不太敏感，可以认为 w_0 是反映 GTT 的基本值. 为方便起见，使用相应的自然周期 $T_0=2\pi/w_0$. 值得注意的事实是，来自各方面的数据表

明，T_0 值小于 4 小时说明情况正常，而明显地大于 4 小时则表示有轻微的糖尿病.

这种简化模型的一个不足是，有时在注入的葡萄糖被吸收后 3～5 小时之间，模型与实际数据拟合不好，这表明，肾上腺素和糖原这样的变量在此期间起了重要作用. 因此，这些变量应作为独立变量包含在模型中.

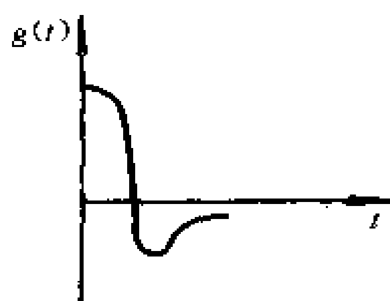


图 5-4　当 $a^2-w_0^2>0$ 时，$g(t)$ 的曲线图

在(5)式中，若 $a^2-w_0^2>0$，则 $g(t)$ 有如下形式(如图 5-4).

注意到 $g(t)$ 从最高点迅速下降，体内分泌大量的内分泌激素，这种情况往往表示患有急症.

在模型的讨论中看到，对内分泌激素采用集中参数方法简化模型，在诊断糖尿病上基本上是成功的. 但若对内分泌激素分别考虑，即采用分布参数，则模型将会更准确.

§5.5　弱肉强食模型

在生态系统中，没有一种动物是完全孤立生活的. 因为所有的动物都要通过食物来生活，它们一定相互影响. 如果不吃其他动物，也要吃植物. 在这里考虑一个两种群的系统，一种以另一种为食，比如狐狸(捕食者)与兔子(被捕食者)，这种系统称为"寄主——寄生物"系统或"被食者——捕食者"系统，或借用替代说法"食草者——肉食者"系统.

该模型的研究有一段历史背景. 本世纪 20 年代意大利生物学家 D'Ancona 研究了各种鱼类的相互制约关系，他从第一次世界大战期间地中海各港口捕获的各种鱼类占总捕获量的百分比资料

中发现：鲨鱼等掠肉鱼(以掠食用鱼为生的鱼)的比例在战争期间明显地增加. 他知道，捕获的各种鱼的比例基本上可以代表地中海渔场中各种鱼的比例. 战争使捕获量下降，渔场中食用鱼增加，以此为生的鲨鱼等掠肉鱼也随之增加，但是为什么捕获量的下降会使鲨鱼等的比例增加，即对鲨鱼的生存更有利呢？他无法解释这种现象，于是请 V. Volterra 建立数学模型，回答上述问题.

Volterra 提出：记食用鱼数量为 $x(t)$，鲨鱼等掠肉鱼的数量为 $y(t)$，因为大海的资源很丰富，可以认为如果 $y(t)=0$，则 $x(t)$ 将以自然增长率 $r(r>0)$增长，即 $\dot{x}=rx$. 但是鲨鱼以食用鱼为食，致使鱼的增长率降低，设降低程度与鲨鱼数量 $y(t)$成正比，于是相对增长率为 $r_x=r-\lambda y$. 常数 $\lambda>0$，反映了鲨鱼掠取食用鱼的能力. 如果没有食用鱼，鲨鱼无法生存，设鲨鱼的自然死亡率为 d，则 $\dot{y}=-dy$. 食用鱼为鲨鱼提供了食物，致使鲨鱼死亡率降低，或者说为鲨鱼提供了增长的条件. 设增长率与食用鱼的数量 $x(t)$成正比，于是鲨鱼的相对增长率 $r_y=-d+\mu x$. 常数 $\mu>0$，反映了食用鱼对鲨鱼的供养能力. 所以模型可归结为

$$\begin{cases}\dot{x}=x(r-\lambda y)\\ \dot{y}=y(-d+\mu x)\end{cases}\tag{1}$$

对上述非线性微分方程组作稳定性分析. 首先求平衡点. 令 $\dot{x}=0,\dot{y}=0$，得$(0,0)$和$\left(\frac{d}{\mu},\frac{r}{\lambda}\right)$. 只考虑$\left(\frac{d}{\mu},\frac{r}{\lambda}\right)$的稳定性. 为此作变量代换使原点为平稳点. 令

$$x(t)=\frac{d}{\mu}+\rho(t),\quad y(t)=\frac{r}{\lambda}+\varphi(t)$$

代入(1)式得

$$\begin{cases}\dfrac{\mathrm{d}\rho}{\mathrm{d}t}=-\lambda\varphi\left(\dfrac{d}{\mu}+\rho\right)\\ \dfrac{\mathrm{d}\varphi}{\mathrm{d}t}=\mu\rho\left(\dfrac{r}{\lambda}+\varphi\right)\end{cases}$$

设$|\rho|\ll\frac{d}{\mu}$，$|\varphi|\ll\frac{r}{\lambda}$，去掉二次项使之线性化，得

$$\begin{cases}\dfrac{\mathrm{d}\rho}{\mathrm{d}t}=-\dfrac{\lambda d}{\mu}\varphi\\ \dfrac{\mathrm{d}\varphi}{\mathrm{d}t}=\dfrac{\mu r}{\lambda}\rho\end{cases}$$

即

$$\begin{bmatrix}\dfrac{\mathrm{d}\rho}{\mathrm{d}t}\\ \dfrac{\mathrm{d}\varphi}{\mathrm{d}t}\end{bmatrix}=\begin{bmatrix}0 & -\dfrac{\lambda d}{\mu}\\ \dfrac{r\mu}{\lambda} & 0\end{bmatrix}\begin{bmatrix}\rho\\ \varphi\end{bmatrix}$$

所以

$$\left|\lambda^1 I-\begin{pmatrix}0 & -\dfrac{\lambda d}{\mu}\\ \dfrac{r\mu}{\lambda} & 0\end{pmatrix}\right|=\begin{vmatrix}\lambda^1 & \dfrac{\lambda d}{\mu}\\ -\dfrac{r\mu}{\lambda} & \lambda^1\end{vmatrix}=\lambda^{12}+rd=0$$

$$\lambda^1_{1,2}=\pm\sqrt{rd}i.$$

线性方程组的通解为

$$\rho(t)=\rho_0\cos\sqrt{rd}t,\quad \varphi(t)=\varphi_0\sin\sqrt{rd}t.$$

由此可以看出：

1° 对 ρ 和 φ 的线性化解，既不增长也不衰减，而是连续振动. 这意味着平稳是亚稳定的，这是一种广义稳定(Kolmogorov 意义下)；在平衡点邻域显示出稳定的有界循环.

2° 令 $\sqrt{rd}T=2\pi$，则振动周期 $T=2\pi/\sqrt{\pi d}$.

3° 食用鱼和鲨鱼的总数的振动相位差 90°，食用鱼导前，如图 5-5 所示.

4° 从解中消去时间 t，得

$$\left(\frac{\rho}{\rho_0}\right)^2+\left(\frac{\varphi}{\varphi_0}\right)^2=1$$

这条轨线在总体相空间的图形如图 5-6 所示.

5° 可定出在相空间运动的方向. 由线性化方程

$$\begin{cases}\dfrac{\mathrm{d}\rho}{\mathrm{d}t}=-\dfrac{\lambda d}{\mu}\varphi\\[2mm]\dfrac{\mathrm{d}\varphi}{\mathrm{d}t}=\dfrac{\mu r}{\lambda}\rho\end{cases}$$

可知

$$\begin{cases}\dfrac{\mathrm{d}x}{\mathrm{d}t}=-\dfrac{\lambda d}{\mu}(y-\dfrac{r}{\lambda});\\[2mm]\dfrac{\mathrm{d}y}{\mathrm{d}t}=\dfrac{\mu r}{\lambda}(x-\dfrac{d}{\mu}),\end{cases}$$

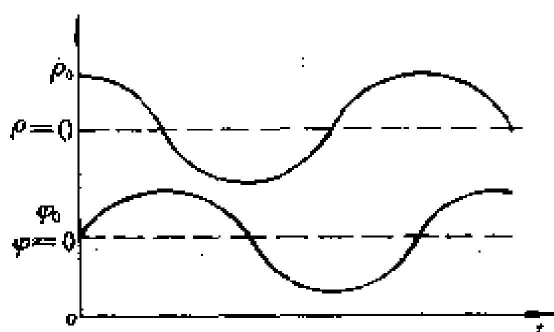


图 5-5　食用鱼和鲨鱼数量变化图示

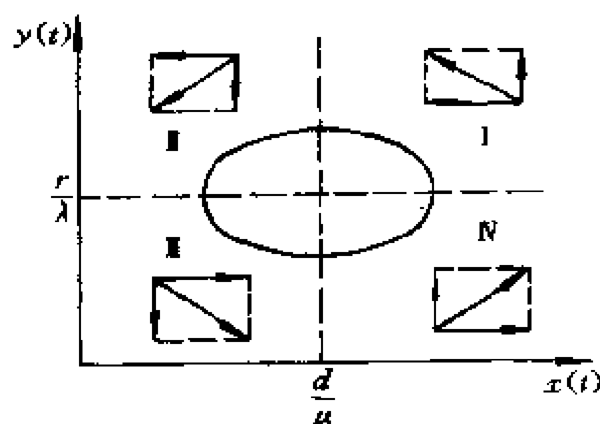


图 5-6　轨线在相空间的图形

因此在图 5-6 所示四个区域中有 Ⅰ：$\dot{x}<0,\dot{y}>0$；Ⅱ：$\dot{x}<0,\dot{y}<0$；Ⅲ：$\dot{x}>0,\dot{y}<0$；Ⅳ：$\dot{x}>0,\dot{y}>0$. 在四个区域中轨线运动方向如图 5-6 所示.

从对非线性微分方程组的直接讨论也可知道轨线是一族以平衡点为中心越来越扩展的封闭曲线. 封闭曲线对应着方程 $\dot{x}=x(x-\lambda y)$，$\dot{y}=y(-d+\mu x)$的周期解. 记周期为 T，$\bar{x}$ 和 $\bar{y}$ 分别为在一个周期内 $x(t)$和 $y(t)$的平均值，则

$$\bar{x}=\frac{1}{T}\int_0^T x(t)\mathrm{d}t=\frac{1}{T}\int_0^T\frac{1}{\mu}(d+\frac{\dot{y}}{y})\mathrm{d}t=\frac{d}{\mu}$$

其中，利用了 $y(t)$为周期函数 $y(T)=y(0)$.

同理，$\bar{y}=\dfrac{r}{\lambda}$.

对于周期性变化的 $x(t)$和 $y(t)$，用它们的平均值 $\bar{x}$ 和 $\bar{y}$ 表示

其大小. 上面的分析表明，食用鱼的(平均)数量取决于鲨鱼方程 $\dot{y}=y(-d+\mu x)$ 中的参数 d 和 μ，而鲨鱼的(平均)数量取决于食用鱼方程 $\dot{x}=x(r-\lambda y)$ 中的参数 r 和 λ. 当鱼的自然增长率 r 下降时，鲨鱼的数量将减少，而当鲨鱼掠取食用鱼的能力提高时，对食用鱼没有影响，只使鲨鱼减少. 另一方面，鲨鱼死亡率 d 上升将导致食用鱼增多，而食用鱼对鲨鱼的供养能力 μ 的提高，会导致食用鱼减少.

上面讨论的是在自然环境中食用鱼和鲨鱼数量的变化规律. 在人工捕捞的情况下，设捕捞能力为 e，则食用鱼的自然增长率由 r 减少为 $r-e$，鲨鱼的死亡率由 d 增加到 $d+e$，在这种变化下，上述模型仍然成立. 于是食用鱼和鲨鱼的(平均)数量变为

$$\bar{x}'=\frac{d+e}{\mu},\quad \bar{y}'=\frac{r-e}{\lambda}$$

相应的平衡点由 p_0 移动到 p'(见图 5-7). 在战争时期捕捞系数从 e

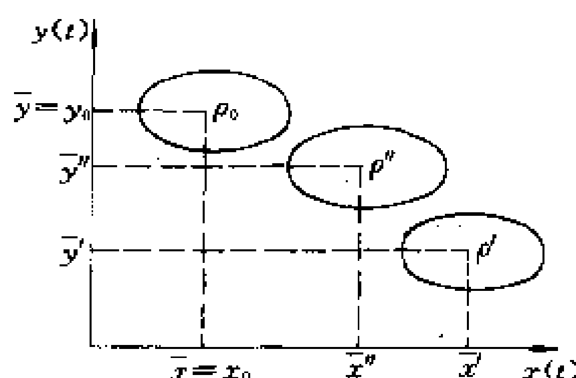


图 5-7 食用鱼和鲨鱼平均数量变化图示

减小到 e_1，食用鱼和鲨鱼的(平均)数量又由 $\bar{x}'$，$\bar{y}'$ 变为

$$\bar{x}''=\frac{d+e_1}{\mu},\quad \bar{y}''=\frac{r-e_1}{\lambda}$$

平衡点由 p' 移动到 p''(见图 5-7). 因为 $e_1<e$，显然 $\bar{x}''<\bar{x}'$，$\bar{y}''>\bar{y}'$. 即捕捞能力的下降会增加鲨鱼的数量而减少食用鱼的数量.

上述模型称为 Lotko-Volterra 模型，它最显著特点是亚稳定性(或中间稳定性，它最早是由 Kolmogorov 于 1936 年给出的)，

然而在此模型中，被食者和捕食者都经历恒定的振动，其振幅与两者的生物性质毫无关系，只取决于完全任意的种群初值大小. 模型的这种“不自然”行为，推动了模型的改进，Leslie 和 Gower1960年考虑了两个种群相对大小对两者增长率的可能影响，并进一步注意到被捕食者中密度相关的限制，其结果是状态空间的轨线以螺旋线形式收敛于平衡点，所以每个种群随时间推移有着衰减的振动而趋于平衡. Holling(1965)-Tanner(1975)考虑了被食者的密度对捕食者的侵害率有可能的影响，发现两个种群的大小对时间而言不断振动，其振幅和周期很快趋于与它们的初始大小无关而只取决模型参数(即方程系数)的极限值. 这种连续的寄主——寄生物(或被食者——捕食者)的循环已被认为非常相似于自然界中许多这类两种群关系的观察. 这一问题利用数据分析法建立起来的模型，参考第八章 § 8.5 加拿大山猫问题.

习　题

1. 一角度高为 60°的圆锥形漏斗里装着 10cm 高的水，其下端小孔的面积为 0.5cm²，求这些水流完需用多少秒？

2. 某容器温度为 60℃，将其内的温度计移入另一容器，十分钟后读数为 70℃，又过十分钟后读数为 76℃，问另一容器的温度为多少度？(提示：牛顿定律指出，温度变化率与温差成正比)

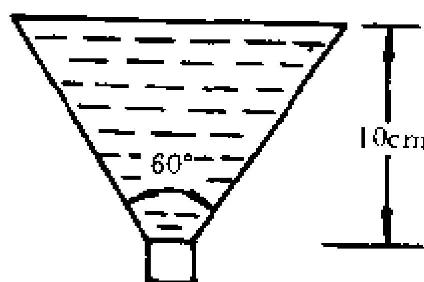


图 5-8　圆锥形漏斗图示

3. 某人每天由饮食获取 10467 焦热量，其中 5038 焦用于新陈代谢，此外每公斤体重需支付 69 焦热量作为运动消耗，其余热量则转化为脂肪. 已知

以脂肪形式贮存的热量利用率为100%，每公斤脂肪含热量41868焦，问此人的体重如何随时间而变化？

4. 在化学反应中，两种物质c_1和c_2等量结合生成新物质c_3. 设a和b是c_1和c_2的初始浓度($t=0$)，定义$x(t)$为c_3在时刻e的浓度，c_3浓度增加速度是与t时刻没有参加反应的c_1和c_2的浓度成正比. 试构造一数学模型.

5. 设某城市共有$n+1$人，其中一人出于某种目的编造了一个谣言. 该城市具有初中以上文化程度的人占总人数一半，这些人只有$\frac{1}{4}$相信这一谣言，而其他人约有$\frac{1}{3}$会相信. 又设凡相信此谣言的人每天在单位时间内传播的平均人数正比于当时尚未听说此谣言的人数，而不相信此谣言的人不传播谣言. 试建立一个反映谣言传播情况的微分方程模型.

6. 取几只容量相等的杯子，充满水，并且一只放在另一只下面排列. 向第一只(即最高位置)杯子以定常速度倒入与杯子等容量的葡萄酒，溢出液体正好流到第二只杯子，第二只杯溢出液体流到第三只杯子…. 假设水和葡萄酒的混合是瞬时发生的，求在任一时刻t及过程结束时T，每只杯子所含葡萄酒量.

7. 建立耐用消费的市场销售量的模型. 如果已知了过去若干时期销售量的情况，如何确定模型的参数.

8. 人工肾是帮助人体从血液中带走废物的装置. 它通过一层薄膜与需要带走废物的血管相通(如图5-9). 人工肾中通以某种液体，其流动方向与血液在血管中的流动方向相反，血液中的废物透过薄膜进入人工肾.

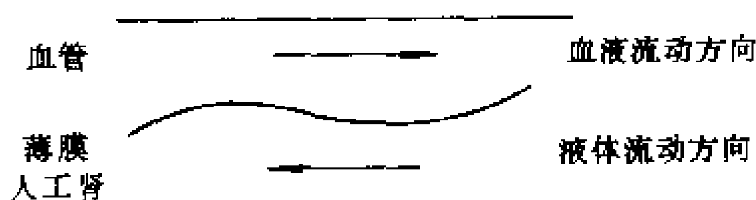


图5-9 人工肾图示

设血液和人工肾中液体的流速均为常数，废物进入人工肾的数量与它在这两种液体中的浓度差成正比. 人工肾总长L. 建立单位时间内人工肾带走废物数量的模型.

9. 根据经验当一种新商品投入市场后，随着人们对它的拥有量的增加，

其销售量 $s(t)$ 的下降速度与 $s(t)$ 成正比. 广告宣传可给销量添加一个增长速度，它与广告费 $a(t)$ 成正比. 但广告只能影响这种商品在市场上尚未饱和的部分（设饱和量为 M）. 建立销量 $s(t)$ 的模型. 若广告宣传只进行有限时间 τ，且广告费为常数 a，问 $s(t)$ 如何变化.

10. 对于技术革新的推广，在下列几种情况下分别建立模型.

(1)推广工作通过已经采用新技术的人进行，推广速度与已采用新技术的人数成正比，推广是无限的.

(2)总人数有限. 因而推广速度随着尚未采用新技术人数的减少而降低.

(3)在(2)的前提下还要考虑广告等媒介的传播作用.

11. 药物进入机体后，在随血液输送到各个器官和组织的过程中，不断被吸收、分布、代谢，最终排出体外. 药物在血液中的浓度，即单位体积血液（毫升）中药物含量（毫克或微克）称为血药浓度，它随时间和空间（机体的各部分）而变化. 研究药物在体内吸、分布和排除的动态过程，及这些过程与药理反应时间的定量关系，是药物动力学的主要内容. 其重要步骤是建立房室模型(Compartment Models). 所谓房室是指肌体的一部分，药物在一个房间内呈均匀分布，即血药浓度为常数，而在不同房室之间按一定规律进行药物的转移. 这种简化假设已由临床试验证明是正确的，并为医学界和药理学界所接受. 图 5-10 所示是一个二室模型.

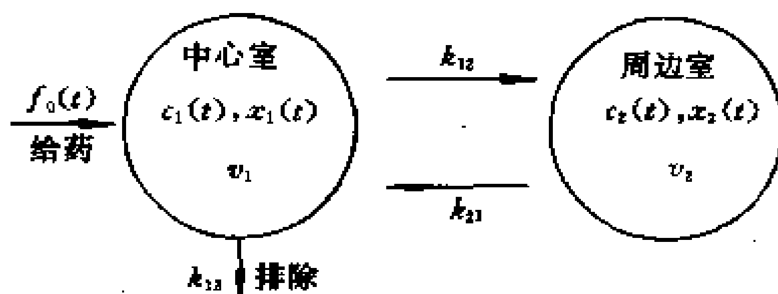


图 5-10　药物动力学的二室模型

其中 $c_i(t)$，$x_i(t)$ 和 v_i 分别表示第 i 室 $(i=1,2)$ 的血药浓度、药量和容积. k_{12} 和 k_{21} 是两室之间药物转速度系数，k_{13} 是药物从中心室向体外排除的速率系数. $f_0(H)$ 是给药速率. 设转移速率及排除速率与该室的血药浓度成正比，与转移和排除的药量相比，药物的吸收可以忽略.

(1)建立两个房室中药量 $x_1(t)$ 和 $x_2(t)$ 满足的微分方程.

(2)讨论在快速静脉注射方式下，动态过程的特性（提示：设在 $t=0$ 时瞬

时将剂量 D_0 的药物输入中心室，则 $f_0(t)$ 和初始条件为：$f_0(t)=0, c_1(0)=\frac{D_0}{V_1}, c_2(0)=0$).

(3)在恒速静脉滴注情况下，动态过程的特性(提示：$f_0(t)=k_0, c_1(0)=0, c_2(0)=0$).

12. 分析具有共生或竞争关系的两种群的增长规律.

第六章　偏微分方程

当用微观来观察现象或过程，而因变量与两个或两个以上自变量有关时，因变量对自变量的变化率导致了偏导数的出现，因此其内在规律的完整描述必须用偏微分方程模型来表示．从建模的角度看，偏微分方程模型比常微分方程模型更为“自然”．偏微分方程模型的建立主要依赖于先验知识，即依赖于机理分析．由于各学科的基本物理定律都满足守恒原理和连续原理．可以在此基础上建立起描述各种物理现象的偏微分方程模型。

§6.1　方程的建立

按所代表的物理过程，方程一般可分为三类：

1° 振动与波（机械的、电磁的）——波动方程．例如，在各向同性的固体中传播的横波或者纵波的方程

$$\nabla^2 u-\frac{1}{a^2}\frac{\partial^2 u}{\partial t^2}=0$$

其中，$u=u(x,y,z,t)$代表平衡时坐标为(x,y,z)的点在t时刻的横向或者纵向位移，a是波的传播速度，∇^2是拉普拉斯算子：

$$\nabla^2=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$$

2° 输运过程——扩散方程、热传导方程．

3° 稳定（或者静止、平衡）过程（或者状态）——拉普拉斯方程．

$$\nabla^2 u=0$$

此外还有量子力学中的波动方程．这种方程所描写的物理过程（或状态）既有波动的特征，又包含统计规律性．

下面举几个例子说明如何从一个物理问题导出偏微分方程．

1．杆的纵振动

考虑一均匀细杆在沿杆长的方向 x 作小振动．令 $u(x,t)$ 表示在平衡时坐标为 x 的点（实际是一截面）在 t 时刻的位移（沿 x 方向）．取不包含端点的一小段 $(x,x+\Delta x)$（如图 6-1）来研究它在 t 时刻的运动．

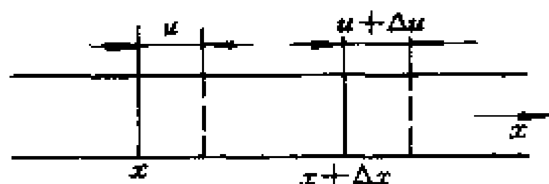


图 6-1　均匀细杆沿杆长方向的振动

以 $\overline{u}(x,t)$ 表示这一小段的质心的位移，则有运动方程

$$\rho s\Delta x\frac{\partial^2\overline{u}}{\partial t^2}=[P(x+\Delta x,t)-P(x,t)]S$$

其中 ρ 是杆的密度，S 是截面积，$P(x,t)$ 是在 x 点的截面上的应力（沿 x 方向）．令 $\Delta x\to 0$，则 $\overline{u}\to u$，

$$\frac{P(x+\Delta x,t)-P(x,t)}{\Delta x}\to\frac{\partial P}{\partial x}$$

则有

$$\rho\frac{\partial^2 u}{\partial t^2}=\frac{\partial P}{\partial x}$$

根据胡克(Hooke)定律，如果略去垂直于杆长方向的形变（对于很细的杆，这是合理的）．则应力 P 与相对伸长 $\frac{\partial u}{\partial x}$ 成正比：

$$P=E\frac{\partial u}{\partial x}$$

其中 E 是杨氏模量，故有

$$\rho\frac{\partial^2 u}{\partial t^2}=\frac{\partial}{\partial x}\left(E\frac{\partial u}{\partial x}\right)=E\frac{\partial^2 u}{\partial x^2}$$

或

$$\frac{\partial^2 u}{\partial t^2}-a^2\frac{\partial^2 u}{\partial x^2}=0$$

其中 $a=\sqrt{\frac{E}{\rho}}$ 是一个常数．

在上面的推导中，我们作了一些简化假定．其一是假定杆只作小振动，这样才能应用胡克定律，并略去由于杆的伸缩所引起的密度 ρ 的变化，而得到一个线性方程．其二是假定杆很细，因此在任何时刻每一横截面上的各点位移相同，而可以用一个变量 x 来标志同一横截面上的各个点，否则 u 将不只是 x 和 t 的函数．这两个假定也指出了该数学模型在应用时受到的限制．

2. 弦的横振动

假设弦是均匀的，并且是完全柔软的，平衡时沿着一条直线绷紧．取这条直线为 x 轴，而以坐标 x 标志弦的各点．设弦在一个平面上振动，以 $u(x,t)$ 表示弦的 x 点在 t 时刻沿垂直于 x 方向的位移．考虑弦的一小段 $(x,x+\Delta x)$ 的运动（如图 6-2）.

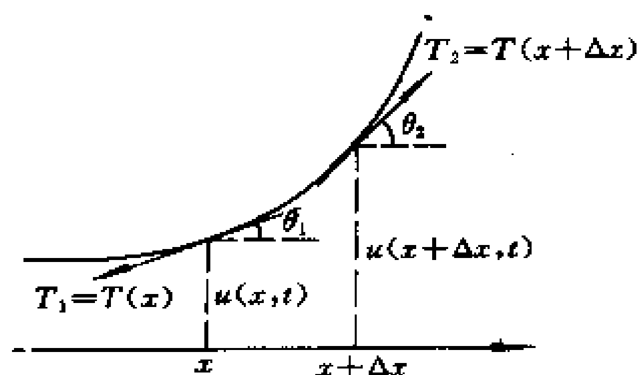


图 6-2　弦的横振动

设 ρ 为弦的单位长质量，则运动方程为

$$\rho\Delta x\,\frac{\partial^2 \bar{u}}{\partial t^2}=T_2\sin\theta_2-T_1\sin\theta_1$$

$$0=T_2\cos\theta_2-T_1\cos\theta_1$$

其中，T_1 和 T_2 分别表示在 x 点和 $x+\Delta x$ 点的张力，θ_1 和 θ_2 为相应倾角．由于假设弦是完全柔软的，故张力沿弦的切线方向．上面第二个方程表示在 x 方向弦是平衡的，因为假定弦只作横振动．

对于小振动，可设 θ 角很小，略去 θ^2，有 $\cos\theta_1\approx\cos\theta_2$．于是由第二个方程得 $T_1=T_2=T$，即在没有沿 x 方向的外力时，弦中各

点的张力是相同的．第一个方程成为

$$\rho\Delta x\frac{\partial^2\bar{u}}{\partial t^2}=T(\sin\theta_2-\sin\theta_1)\approx T(\mathrm{tg}\theta_2-\mathrm{tg}\theta_1)$$

$$=T\left[\left(\frac{\partial u}{\partial x}\right)_{x+\Delta x}-\left(\frac{\partial u}{\partial x}\right)_x\right]=T\frac{\partial^2 u}{\partial x^2}\Delta x,$$

消去两边的 Δx，然后令 $\Delta x\to 0$，则 $\bar{u}\to u$，有

$$\rho\frac{\partial^2 u}{\partial t^2}=T\frac{\partial^2 u}{\partial x^2}$$

或

$$\frac{\partial^2 u}{\partial t^2}-a^2\frac{\partial^2 u}{\partial x^2}=0,$$

其中 $a=\sqrt{\dfrac{T}{\rho}}$.

由于假定 θ 角很小，$\left|\dfrac{\partial u}{\partial x}\right|\ll 1$，若略去 $(\dfrac{\partial u}{\partial x})^2$，则弦的伸长可略去．令 ΔS 表示 $(x,x+\Delta x)$ 这一段的伸长，则

$$\Delta S=\int_x^{x+\Delta x}\sqrt{\mathrm{d}u^2+\mathrm{d}x^2}-\Delta x=\int_x^{x+\Delta x}\sqrt{1+(\frac{\partial u}{\partial x})^2}\mathrm{d}x-\Delta x$$

$$\approx\int_x^{x+\Delta x}\mathrm{d}x-\Delta x=0.$$

因此，弦中张力在任何时刻都是一样的，而 a 为一常数．

在上面的推导中作了一系列的假定．例如，设弦是完全柔软的，只有在这个假定下，张力 T 才总是沿着弦的切线方向（否则得到一个比较复杂的方程，杆的横振动方程）．又如假定了振动很小，因而 θ（或 $\sin\theta$）可用 $\mathrm{tg}\theta$ 代替，而且伸长可以忽略不计，否则张力 T 将由于各点的伸长不同而依赖于 u，得到的方程将不是线性的．

若弦还受到外力的作用，单位长度所受的力为 $F(x,t)$，方向总是垂直于弦平衡时的方向 x，则方程为

$$\frac{\partial^2 u}{\partial t^2}-a^2\frac{\partial^2 u}{\partial x^2}=f(x,t)$$

其中，$f(x,t)=\dfrac{F(x,t)}{\rho}$ 代表单位质量所受的力．

3. 热传导方程

前两例说明了在弹性固体的振动中如何根据牛顿运动定律和弹性定律导出描述这种振动的偏微分方程．前一定律给出位移随时间的变化与应力的关系，后一定律则把应力与位移随位置的变化联系起来，从而得到关于位移的方程．

在固体的热传导问题中也有两个基本定律，能量守恒与热传导的傅里叶(Fourier)定律．

考虑一根均匀细杆，沿着杆长方向 x 维持一定的温度差(如图 6-3 所示)，实验结果表明，在杆中有热量传导，热量由温度高处"流"向温度低处，单位时间内通过单位面积的热量 q 与温度 u 的下降率成正比：

$$q=k\frac{u-(u+\Delta u)}{\Delta x}=-k\frac{\Delta u}{\Delta x}$$

取极限得

$$q=-k\frac{\partial u}{\partial x}$$

其中 q 称为热流密度，系数 k 称为导热率，导热率与杆的质料有关，而且严格说来也与温度有关．但如果温度的改变范围不太大，k 可以近似地看作常数．这就是傅里叶定律的数学表示．

在空间的热传导，热流密度是矢量，它在各向同性的物体中与温度下降率的关系为

$$q_x=-k\frac{\partial u}{\partial x},q_y=-k\frac{\partial u}{\partial y},q_z=-k\frac{\partial u}{\partial z}$$

或
$$q=-k\nabla u$$

现考虑各向同性固体热传导问题．取物体内部任意一个小长方体 $ABCDEFGH$(如图 6-3 所示)，取坐标系 $o\,xyz$，使坐标面与这长方体的面平行．

在 Δt 时间内沿 x 方向通过与 yz 平面平行的 $CDEH$ 面进入六面体的热量是$(q_x)_x\Delta y\Delta z\Delta t$，而通过 $ABGF$ 面从六面体出去的热量是$(q_x)_{x+\Delta x}\Delta y\Delta z\Delta t$，因此净入热量是

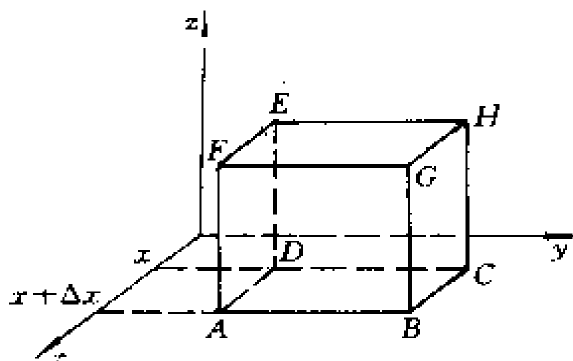


图 6-3　物体内部小长方体

$$(q_x)_x\Delta y\Delta z\Delta t-(q_x)_{x+\Delta x}\Delta y\Delta z\Delta t$$

$$=-\frac{\partial}{\partial x}\left(-k\frac{\partial u}{\partial x}\right)\Delta x\Delta y\Delta z\Delta t=\frac{\partial}{\partial x}\left(k\frac{\partial u}{\partial x}\right)\Delta v\Delta t$$

其中 Δv 是长方体的体积 $\Delta x\Delta y\Delta z$. 同样,在 Δt 时间内沿 y 和 z 方向进入长方体的热量分别是

$$\frac{\partial}{\partial y}\left(k\frac{\partial u}{\partial y}\right)\Delta v\Delta t \text{ 和 } \frac{\partial}{\partial z}\left(k\frac{\partial u}{\partial z}\right)\Delta v\Delta t.$$

进入长方体的热量使其温度升高．设 ρ 为物体的密度,c 为比热,则

$$\rho c\Delta v\Delta u=\left[\frac{\partial}{\partial x}\left(k\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial y}\left(k\frac{\partial u}{\partial y}\right)+\frac{\partial}{\partial z}\left(k\frac{\partial u}{\partial z}\right)\right]\Delta v\Delta t$$

消去 Δv,令 $\Delta t\to 0$,得

$$\rho c\frac{\partial u}{\partial t}=\frac{\partial}{\partial x}\left(k\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial y}\left(k\frac{\partial u}{\partial y}\right)+\frac{\partial}{\partial z}\left(k\frac{\partial u}{\partial z}\right)$$

若物体是均匀的,则 k 与坐标无关,方程简化为

$$\frac{\partial u}{\partial t}-a^2\nabla^2 u=0$$

这是各向均匀同性物体的热传导方程,其中 $a^2=\frac{k}{\rho c}$ 称为扩散率或传导率．

如果在物体中有热产生,设在单位时间内单位体积中产生的热量为 $F(x,y,z,t)$,则

$$\frac{\partial u}{\partial t}-a^2\nabla^2 u=f(x,y,z,t),$$

其中，$f(x,y,z,t)=\dfrac{F(x,y,z,t)}{\rho c}$.

方程中的参数 a^2 一般与温度 u 有关，因为密度 ρ、比热 c 和导热率 k 都与温度有关．但如果温度变化的范围不太大，则 a^2 可近似认为与温度无关而是一个常数．

对于分子扩散问题，方程是相同的，只是 u 代表浓度，系数 a^2 应为扩散率 D，而 f 代表扩散物质的生产率．

在一定条件下，当物体的温度分布达到稳定状态时，$\frac{\partial u}{\partial t}=0$，方程成为稳定温度场的方程

$$\nabla^2 u=-\frac{f}{a^2}$$

称为泊松(Poisson)方程．若 $f\equiv 0$，则称为拉普拉斯方程．

在理论和实际问题的讨论中，为方便起见，需要将偏微分方程分类．

我们知道，在研究二元二次代数方程时，通过变量代换能将其化成哪种标准形式，决定于二次形

$$f(x,y)=a_{11}x^2+2a_{12}xy+a_{22}y^2$$

的代数性质．或者说，由二次曲线 $f(x,y)=1$ 的性质决定，这个曲线可以是一个椭圆，一个双曲线，或者一个抛物线．相应地，定义方程在某一点的类型如下．

如果二阶线性方程

$$a_{11}u_{xx}+2a_{12}u_{xy}+a_{22}u_{yy}+b_1u_x+b_2u_y+cu=f$$

(这里 $a_{11},a_{12},a_{22},b_1,b_2,c,f$ 都是变量 x,y 在某一区域 Ω 上的连续可微实函数)中的二阶导数项的系数在区域 Ω 中某点 (x_0,y_0) 满足

$$\Delta\equiv a_{12}^2-a_{11}a_{22}>0$$

则称方程在点 (x_0,y_0) 为双曲型的；若在点 (x_0,y_0) 满足

$$\Delta\equiv a_{12}^2-a_{11}a_{22}=0,$$

则称方程在点 (x_0,y_0) 为抛物型的；若在点 (x_0,y_0) 满足

$$\Delta \equiv a_{12}^2 - a_{11}a_{22} < 0$$

则称方程在点(x_0, y_0)为椭圆型的.

其次,如果方程在一个区域Ω中的每点均为双曲型,那么就称方程在区域Ω内是双曲型的.类似可定义方程在Ω内是抛物型或椭圆型.有些方程在区域Ω的一部分内是双曲型的,而在另一部分是椭圆型的,在它们的分界线上是抛物型的,称为在区域Ω中是混合型的.

§6.2 边界条件和初值条件

前面讨论了如何从一个物理问题和相应的物理定律中建立起它的数学模型——偏微分方程.然而仅有方程还不足以确定物体的运动,因为物体的运动还与初始状态以及通过边界所受到的外界作用有关.这一点是连续体(或者场)运动的特点,即外界的作用常常是通过物体的边界"传"到物体内部去的.从数学的角度看,一个微分方程有无穷多的解,表现在其通解中含有若干个任意常数或者任意函数,而初始状态和边界情况则是确定这些常数的数值或者函数的形式的初值条件和边界条件.求一个微分方程的解满足一定初值条件和边界条件的问题称为定解问题.

某个物理过程在一定条件下总是具有唯一确定的状态,因此正确描述这种物理过程的定解问题的解应是存在唯一的.但是用偏微分方程对此作出描述时,总要经过一些近似处理(例如要舍弃一些因素)并提出一些附加条件,只能说定解问题近似地反映了自然现象中某个物理过程,而不能说两者完全等同.如果定解问题的解存在而且唯一,称定解问题是适定的.因此,必须判断所给定的条件(初值或边值)是否足以保证解的存在和唯一性.另外,出现在定解条件中的函数值通常是由实验测量确定的,一般只是近似的.在这种定解条件下求得的解,只有在它和准确的定解条件下确定的解相差很小时,才有实际意义.如果定解条件的微小变

化引起的解在定义域中的变化也是微小的，则称定解问题是稳定的，或称解是连续依赖于定解条件的.

确定定解条件和推导偏微分方程一样也是数学建模的重要步骤，适定性的研究涉及到很多较深的理论问题，有兴趣的读者可参考有关数理方程的专著. 这里仅将几类偏微分方程的常用定解条件列出，为建模提供方便.

考虑最简单的抛物型方程

$$\frac{\partial u}{\partial t}-a^2\frac{\partial^2 u}{\partial x^2}=f(x,t),a>0,(x,t)\in D \tag{1}$$

其中，D 是 x-t 平面内的给定区域，它可以是有界区域，也可以是无界区域.

方程的定解问题根据定解条件可分为四种类型.

1° 初值问题(Cauchy 问题，也称自由边界问题)

在区域 $D=\{(x,t)|x\in(-\infty,+\infty),t>0\}$ 内求方程(1)满足初始条件

$$u=(x,0)=\varphi(x),x\in(-\infty,+\infty)$$

的解 $u(x,t)$.

2° 第一类初边值问题

在区域 $D=\{(x,t)|x\in(0,l),t\in(0,T)\}$ 内求方程(1)满足初值条件和第一类边值条件

$$u(x,0)=\varphi(x),x\in(0,l)$$

$$u(0,t)=\alpha_1(t),u(l,t)=\alpha_2(t),t\in(0,T)$$

的解 $u(x,t)$.

3° 第二类初边值问题

在区域 $D=\{(x,t)|x\in(0,l),t\in(0,T)\}$ 内求方程(1)满足初始条件和第二类边值条件

$$u(x,0)=\varphi(x),x\in(0,l)$$

$$\frac{\partial u}{\partial x}(0,t)=\beta_1(t),\frac{\partial u}{\partial x}(l,t)=\beta_2(t),t\in(0,T)$$

的解 $u(x,y)$.

4° 第三类初边值问题

在区域 $D=\{(x,t)\,|\,x\in(0,l),t\in(0,T)\}$ 内求方程(1)满足初始条件和第三类边值条件

$$u(x,0)=\varphi(x),x\in(0,l)$$

$$u-h_1(t)\frac{\partial u}{\partial x}\Big|_{x=0}=r_1(t),$$

$$u-h_2(t)\frac{\partial u}{\partial x}\Big|_{x=l}=r_2(t),t\in(0,T)$$

的解 $u(x,t)$.

第一类边值条件反映了边界上的温度为已知

$$u|_\Sigma=\varphi(M,t)$$

其中,M 代表区域的边界 E 上的变点,φ 是已知函数.

第二类边值条件是在边界上流入的热流密度为已知:

$$-q_n|_\Sigma=\psi(M,t)$$

其中,q_n 代表热流密度在边界面的外法线方向的分量,ψ 为已知函数,据 Fourier 定律 $q=-k\nabla u$,故 $q_n=-k\dfrac{\partial u}{\partial n}$,因而

$$\frac{\partial u}{\partial n}|_\Sigma=\frac{1}{k}\psi(M,t).$$

如果边界是绝热的,则 $\psi=0$.

第三类边值条件反映了物体表面与外界通过辐射或对流过程交换热量,单位时间内单位表面积失去的热量与表面和外界的温度差成正比.

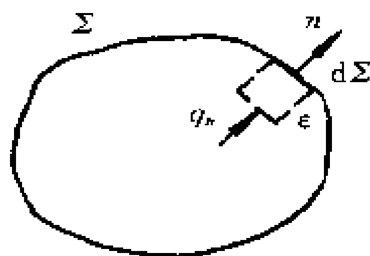


图 6-4 物体内边界处的小柱体

如图 6-4 所示,考虑物体内以边界面上的面积元恒为底,高为 ε 的一个小柱体.在时间 dt 内由物体内部沿$d\Sigma$的外法线方向进入小柱体内的热量是$q_n d\Sigma dt$,而由于与外界交

换热量而失去的热量是$H(u|_\Sigma-u_0)\mathrm{d}\Sigma\mathrm{d}t$. H 是常数 u_0 为外界温度．从小柱体的侧面进出的热量以及在小柱体内由于温度升降的内能改变($=\rho c\mathrm{d}\Sigma\cdot\epsilon\Delta u$)都与 ϵ 成比例，当 $\epsilon\to0$ 时，这两项都趋于零，因此有

$$q_n|_\Sigma=H(u|_\Sigma-u_0)$$

再据 Fourier 定律，得

$$-k\frac{\partial u}{\partial n}\Big|_\Sigma=Hu|_\Sigma-Hu_0$$

令 $h=\frac{H}{k}$，有

$$\left[\frac{\partial u}{\partial n}+hu\right]_\Sigma=hu_0$$

上述三种形式的边界条件是很普遍的边界条件，不限于热传导问题，因此在数学物理中常把它们称为第一、二、三类边界条件．第一类边界条件给出未知函数在边界上的值；第二类边界条件给出未知函数在边界面的法线方向的导数之值；第三类边界条件给出的则是未知函数和它的法线方向导数之间的一个线性关系．可能出现这种情况，在边界的不同部分上边界条件的类型不同．

除了上面列举的三种边界条件之外，还有各式各样的边界条件．例如，在热传导问题中有辐射条件：

$$-\frac{\partial u}{\partial n}\bigg|_\Sigma=c(u^4|_\Sigma-u_0^4)$$

其中，c 是一个常数，u_0 是外界的温度，u 和 u_0 都要用绝对温标．这个边界条件右方含 u 的四次方，故为非线性的．但若 u 和 u_0 相差不大，则 $u^4-u_0^4\approx4u_0^3(u-u_0)$，则转化为第三类边值条件．

连接条件．在两种不同媒质的分界面上有连接条件．例如由两根不同质料的杆接成的一根杆，当杆作纵振动时，在连结点 c 应满足下列条件：

$$u_1|_{x=c}=u_2|_{x=c},\quad E_1\frac{\partial u_1}{\partial x}\Big|_{x=c}=E_2\frac{\partial u_2}{\partial x}\Big|_{x=c}$$

其中，$u_1=u_1(x,t)$，$u_2=u_2(x,t)$分别代表杆的两部分的位移（相对于平衡位置），E_1 和 E_2 分别为两部分的杨氏模量．第一个条件是显然的，因为在连接点 $x=c$，位移必相等，否则杆的两部分便会分开，第二个条件读者可自行证明．

§6.3 场的特性与建模

前面已经讲了如何根据某个物理过程利用微元法建立偏微分方程，也可以利用不同物理过程的相似性，适当选取因变量，建立起形式相同的偏微分方程，仅仅在参数的物理含义上存在差异．

"场"的概念已广泛地用于物理学和工程学．在这些学科中，因变量（表示场的扰动及响应的物理量）被定义为空间和时间的连续函数．各种参数介质的内在性质，也被连续地定义在时间/空间连续系统各点上．由于物理和工程等学科在自变量、因变量和参数之间在概念上的相似性，以及支配着系统特性的各学科的基本定律之间的类似性，使描述分布参数系统的偏微分方程模型可以大为简化，这也使得可以用三类基本偏微分方程及其变形来描述几乎所有物理与工程领域内的数学模型．

偏微分方程的因变量都是物理量，它们构成系统的激励和输出效应．各物理领域中的这些因变量分为两大类．第一类为位势函数，亦称为交叉变量，交叉变量建立了场内某点与其他点（或任意参考点）上条件间的关系．记录交叉变量用的量测仪表，必须同时置于两个不同的点上，测出的变量幅值表示两点的数值差．例如传热系统中的温度，重力系统中的压强，流体力学系统中的压强和势能，静电系统和电动力系统中的压强或电势等等．另一类是流量函数，也称为穿透变量．穿透变量只需要对场内的一点进行测定，其值表示流过场内单元横断面的通量．例如热系统中的热通量、流体力学系统中的流速、电动学系统中的电流密度等等．从数学上讲，交叉变量是标量，而穿透变量是矢量．

根据场内组成物质的固有物理性能，可以导出数学模型中的参数．对提出偏微分方程模型的各物理领域调查以后，不难发现，参数的类型不外乎下列三种：耗散和阻尼 D，位势储蓄 EP 和流量储蓄 E_f．参数 D 是对场的组成物质的能量转换成热量所耗散的度量，或者是在热传递系统中增加熵所耗散的度量，例如电气系统中的电阻率，流体流动系统中的粘度，传热系统中的热阻率等等．当场的组成物质是耗散器时，位势和流量之间的激励/响应关系是一种比例关系．例如电气系统中的一根细线，将位势差 Δp 施加于该系统某小段的两端，进而测得流量 f，则

$$\Delta p = -D\Delta x f$$

其中，D 是单位长度的耗散（电阻率），Δx 是小段的长度，负号表示正流量在位势减少方向流动．若 $\Delta x \to 0$，则

$$\frac{\mathrm{d}p}{\mathrm{d}x} = -Df$$

在多维情况下，有

$$\nabla \boldsymbol{P} = -D\boldsymbol{f}$$

其中，等号左端项被称为位势梯度．

假如场的组成物质能够临时存储物质或能量，则该物质充当一个储蓄器。在位势储蓄器的情况下，每当物质样本上跨接位势差时，便发生存储；位势差越大，存储量越大．例如，每当电容两端跨接电位差时，电容便存储电能，当在两端接入电阻或者短接时，电容将在阻止电位变化的方向上产生电流，流量与位势的变化率成正比．在具有位势储蓄特性的一维场中，位势 $p(x,t)$ 和流量 $f(x,t)$ 的关系为

$$\Delta f = -E_p \Delta x \frac{\mathrm{d}p}{\mathrm{d}t}$$

其中，E_p 是每单位长度的位势储蓄特性．令 $\Delta x = 0$，

$$\frac{\partial f}{\partial x} = -E_p \frac{\partial p}{\partial t}$$

对于多维系统，有

$$\nabla \cdot f = -E_p \frac{\partial p}{\partial t}.$$

分布式电容或电容率是电场内储蓄位势的参数，在流体流动系统中，E_p 是压缩率，在热传递系统中，E_p 是热容量或辐射率．

在流量储蓄器的情况中，每当流量流经场内物质，则该物质便将能量存储下来．例如，在电路中，每当电流流经电感，则电感便存储其磁场的能量．若电感两端短路，则产生电压，其极性是阻碍电流变化的方向，其大小与电流变化率成正比，即

$$\Delta p = -E_f \Delta x \frac{\mathrm{d}f}{\mathrm{d}t}$$

其中，E_f 是单位长度的流量储蓄．令 $\Delta x \to 0$，则

$$\frac{\partial p}{\partial x} = -E_p \frac{\partial f}{\partial t}$$

对于三维情况，有

$$\nabla P = -E_f \frac{\partial f}{\partial t}$$

表 6-1　某些物理领域中的场参数

物理领域	交叉变量 (p)	穿透变量 (f)	耗散参数 (D)	流量储蓄 E_f	位势储蓄 E_p
电动力学	电压	电流	电阻率	电感率	电容率
静电学	电势	电通量	/	/	介质的介电常数
磁学	磁势 Mmf	磁通量	磁阻率	导磁率	/
电磁学	电磁势 EM	电磁通量	导电率	导磁率	介质的介电常数
静力学	位移	力	/	弹性常数	/
动力学	位移或速度	力	粘性阻尼	弹性常数	质量
弹性力学	应变	应力	粘性阻尼	杨氏模量	惯性
流体力学	应变	应力	粘性阻尼	惯性(密度)	压缩率
粒子扩散	浓度	质量传输率	扩散率	惯性力	压缩率
传热学	温度	热通量	热阻	/	热容量

分布电感或电感率是分布电系统中流量储蓄参数，而在流体流动系统中，密度和质量是流量储蓄参数．传热系统不具有流量储蓄器的性质．

某些物理领域中的场参数如表 6-1 所示．

在确定系统性能及其数学描述时，系统的参数的类型起着十分重要的作用．在只有一个参数呈现（即此参数与其他参数相比具有足够大的幅值）的场合，数学模型采用椭圆型方程，诸如拉普拉斯方程或泊松方程；在耗散以及一种储蓄（位势储蓄或流量储蓄）呈现的场合，采用抛物型方程．在两种储蓄类型都呈现的场合，则采用双曲型偏微分方程来表征场特性．

一、椭圆型偏微分方程

椭圆型偏微分方程表征这样一些分布参数系统，其中三种参数形式中至少有一种是不可忽视的，即凡具有纯耗散、纯位势储蓄或纯流量储蓄的场，都可以用椭圆型偏微分方程来描述．另外，它也用来表示这样的场：系统含有耗散以及储蓄元素，但它们都已达到静态或稳态．

1．方程的建立

研究图所示系统：一根细导线的一端与具有恒值位势 p_0 的电源相接，该电线的耗散参数是常值，电线的另一端与零位势电源相接．该系统中 E_p 和 E_f 可以忽略不计，故在边界点 x_1 和 x_2 上，位势/流量关系为

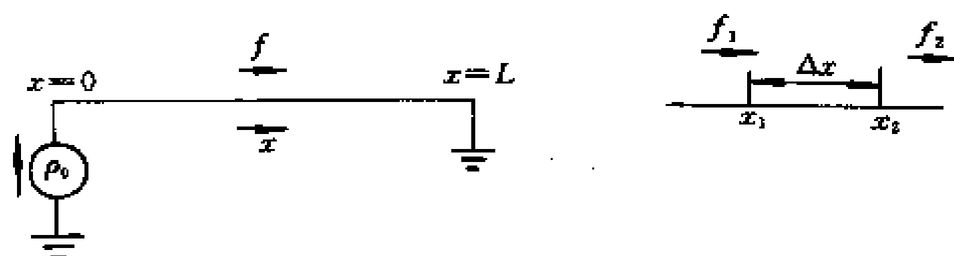


图 6-5　一维电阻场

$$\left(\frac{\mathrm{d}p}{\mathrm{d}x}\right)_1=-Df_1, \quad \left(\frac{\mathrm{d}p}{\mathrm{d}x}\right)_2=-Df_2$$

因为假定场是纯耗散的，在元素段内不可能有存储，按照守恒定律，有

$$f_2-f_1=D$$

故有
$$\frac{1}{D}\left(\frac{\partial p}{\partial x}\right)_2-\frac{1}{D}\left(\frac{\partial p}{\partial x}\right)_1=0$$

令 $\Delta x\to 0$
$$\frac{\partial^2 p}{\partial x^2}=0$$

该方程被称为一维拉普拉斯方程，其边界条件为

$$p=p_0, \text{当 } x=0 \text{ 时}$$

$$p=p_L=0, \text{当 } x=L \text{ 时}$$

对于三维情况，拉普拉斯方程为

$$\frac{\partial^2 p}{\partial x^2}+\frac{\partial^2 p}{\partial y^2}+\frac{\partial^2 p}{\partial z^2}=0$$

在指定的边界条件下，它将完全表征场内的位势分布．

引入拉普拉斯算子∇^2（或 Δ），可将拉普拉斯方程表示为更紧凑的形式

$$\mathrm{div}(\mathrm{grad}\,p)=\nabla\cdot(\nabla p)=0$$

或
$$\nabla^2 p=0$$

边界条件有两种：位势或流量边值．上例是对位势给出的．对于流量，若没有流量经过无激励的边界，则与边界正交的位势梯度为零，若在边界上给定非零流量，则与该边界正交的位势梯度应与流量成比例．该边界条件可以表示成更通用的形式

$$p=k_1, \quad \frac{\partial p}{\partial n}=k_2$$

其中，k_1 和 k_2 是给定常数或空间变量的函数．n 是与边界正交的方向．

2. 方程的修正形式

1° 非均匀性

在上述方程的推导中，假定场特性是均匀而各向同性的，对于 D 是依赖于空间的参数，上述推导应作适当修正.

在一维情况，令 $D=D(x)$，则由

$$f_1=-\frac{1}{D_1}\left(\frac{\mathrm{d}p}{\mathrm{d}x}\right)_1, f_2=-\frac{1}{D_2}\left(\frac{\mathrm{d}p}{\mathrm{d}x}\right)_2$$

有

$$\frac{1}{\Delta x}\left[\frac{1}{D_2}\left(\frac{\mathrm{d}p}{\mathrm{d}x}\right)_2-\frac{1}{D_1}\left(\frac{\mathrm{d}p}{\mathrm{d}x}\right)_1\right]=0$$

令 $\Delta x\to 0$，

$$\frac{\partial}{\partial x}\left(\frac{1}{D(x)}\frac{\partial p}{\partial x}\right)=0$$

对于三维空间的非均匀场，有

$$\frac{\partial}{\partial x}\left(\frac{1}{D}\frac{\partial p}{\partial x}\right)+\frac{\partial}{\partial y}\left(\frac{1}{D}\frac{\partial p}{\partial y}\right)+\frac{\partial}{\partial z}\left(\frac{1}{D}\frac{\partial p}{\partial z}\right)=0$$

上述推导，对 $D=D(p)$ 也适用.

2° 活动坐标

若流量流经的介质本身是移动的场，拉普拉斯方程也要作修正，这种情况相当于坐标系统的原点（空间各位置都以它为基准点进行量测）正以某一指定的速度移动. 例如，在静态或稳态的热传递中，考虑服从拉普拉斯方程的某种现象. 若流体是静止的，则场内温度分布将服从基本拉普拉斯方程 $\nabla^2 p=0$. 若组成场的流体沿 x 方向流动，则热传递不仅受到温度梯度的影响，而且会受到流体粒子速度的影响，对于一维和三维空间情况，分别修正为

$$\frac{\mathrm{d}^2 p}{\mathrm{d}x^2}=\alpha v_x\frac{\mathrm{d}p}{\mathrm{d}x}$$

$$\frac{\partial^2 p}{\partial x^2}+\frac{\partial^2 p}{\partial y^2}+\frac{\partial^2 p}{\partial z^2}=\alpha\left(v_x\frac{\partial p}{\partial x}+v_y\frac{\partial p}{\partial y}+v_z\frac{\partial p}{\partial z}\right)$$

其中 v_x，v_y 和 v_z 分别是 x、y 和 z 方向上坐标系统的速度. 等式右端项正是 v_p 的散度 $\mathrm{div}(vp)$.

3° 源和/或汇集点

当场内有分布的源或流量汇集点起作用时，在一维拉普拉斯方程的推导中应加入流量的输入（或输出）项．

$$f_2 - f_1 = -f_i \Delta x, \quad 得 \quad \frac{\partial^2 p}{\partial x^2} = -f_i$$

其中，f_i 为单位长度上的流入量．

对于非均匀三维问题，拉普拉斯方程的通式为

$$\nabla \cdot \left(\frac{1}{D} \nabla p \right) = -kf_i$$

该方程称为泊松方程．

3. 场特性

拉普拉斯方程描述的场具有某些重要特性：

1° 稳态特性．拉普拉斯方程不包含时间独立变量，这是由于系统内只有一种参数类型直接造成的．由于不存在另外两种参数（位势和流量的储蓄），当激励是时间的函数时，系统各处的响应将具有完全相同的瞬态特性，而且在场内各点上立即达到稳态．

2° 激励仅施加在边界上或场的尽头，在端的内部不出现任何位势源或流量源．因此根据守恒原理，出入场的能量或物质一定是由外部源造成的，而且所有的流量线必定终止在边界上，这就意味着场内不大可能有位势最大的值或最小值．若出现位势最大值，就意味着在场内某些点比邻点有更高的位势．而流量总是从高位势流向低位势，这样，具有极大值的点将向各个方法流出能量或物质．在矢量分析中，称为发散．因而，服从拉普拉斯方程的场一定是没有发散的（散度为零）．同样，位势极小意味着存在能量或物质的汇集点，这也将破坏连续守恒原理，意味着在场内的位势梯度不可能改变极性．

拉普拉斯方程和泊松方程的适用范围很广．

在物理学中有三种基本物理的引力场，即由物质、电荷及磁极产生的场．所有不含这些场干扰的自由空间区域是服从拉普拉斯方程的系统．

在流体流动系统中，若液体粒子流过一种含有细空隙的介质

（如坚实的沙子），则粘滞磨擦加于液体的力将大大超过惯性力．又若液体是不可压缩的，即不可能有液体粒子的存储，则该系统是纯耗散的，可以用拉普拉斯方程来描述．

当不可压缩的液体流入松疏的孔或管道时，通常可以认为它是粘滞性忽略不计的理想流体，在这些情况下，只考虑惯性力（动能），因而该系统也服从拉普拉斯方程．在流体流动系统中，流量函数（穿透变量）是流体粒子的稳定速度，由于流体粒子也可以绕自身的轴而旋转，因此若要使用拉普拉斯方程，就必须是在不旋转的流动情况、对矢量场，拉普拉斯方程描述的调和场，是无源无旋场，即

$$\mathrm{div}\,p=0,\quad \mathrm{vot}\,p=0$$

二、抛物型偏微分方程

抛物型方程在不同场合也被称作扩散方程或传导方程．首先，对简单的一维系统推导其方程，如图 6-6 所示．这时，场内不仅包含分布耗散，而且还有分布储蓄特性（位势或流量）．图示系统可以用来描述一个通过电解质体（位势储蓄）与大地耦合的电阻器（耗散），也可以表示一个呈现热阻（耗散）和热容器（位势储蓄）的热导体．

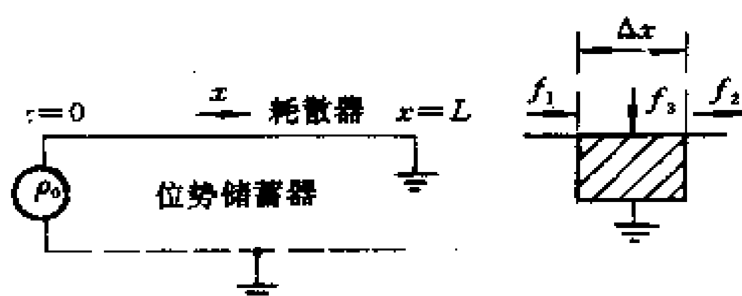


图 6-6　具有位势储蓄器的一维阻抗器

现在除了沿导体的流量 f_1 和 f_2 外，还有进入储蓄器的流量 f_3，根据守恒原理，有

$$f_1-f_2-f_3=0 \tag{1}$$

用 E_p 表示单位长度的存储特性，则

$$f_3 = E_p \Delta x \frac{\partial p}{\partial t} \tag{2}$$

其中，p 是微分元素相对于大地的平均位势，$E_p\Delta x$ 是该元素的总储蓄容量．

对流量 $f_1 - f_2$ 利用 Taglor 级数展开，忽略高阶项，则

$$f_1 - f_2 = -\frac{\partial f}{\partial x}\Delta x \tag{3}$$

又因为流量与位势梯度的关系为

$$f = -\frac{1}{D}\frac{\partial p}{\partial x}$$

对 x 微分，得

$$-\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(\frac{1}{D}\frac{\partial p}{\partial x}) \tag{4}$$

将(3)式代入(2)式，得

$$-\frac{\partial f}{\partial x}\Delta x = E_p \Delta x \frac{\partial p}{\partial t} \tag{5}$$

将(4)式代入(5)式，得

$$\frac{\partial}{\partial x}\left(\frac{1}{D}\frac{\partial p}{\partial x}\right) = E_p \frac{\partial p}{\partial t}$$

该方程称为一维扩散方程．由于方程含有时间 t 和空间位置变量，所以在求解时，还要求提供初始时刻 $t=0$ 时，场上各点的位势 $p(x,0)$（初值条件）和边值条件：$p(0,t)=p_0(t)$，$p(L,t)=p_2(t)$．这时便可确定 $t>0$ 时场内各点的位势．

对于三维扩散场，有

$$\frac{\partial}{\partial x}\left(\frac{1}{D}\frac{\partial p}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{1}{D}\frac{\partial p}{\partial y}\right) + \frac{\partial}{\partial z}\left(\frac{1}{D}\frac{\partial p}{\partial z}\right) = E_p \frac{\partial p}{\partial t}$$

对于均匀 D，有

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = DE_p \frac{\partial p}{\partial t} \quad 或 \quad \nabla^2 p = k\frac{\partial p}{\partial t}$$

当然还应包括相应初边值条件．其中，对初值有两种提法：若有位势储蓄存在，则必须给定位势分布的初值 $p(x,y,z,0)$；若有流量

储蓄存在，则必须给定位势变化率的初值$\frac{\partial p(x,y,z,0)}{\partial t}$.

当发生扩散的介质本身也是运动的，即坐标系是运动的，则扩散方程为

$$\nabla\cdot\left(\frac{1}{D}\nabla p\right)=E_p\frac{\partial p}{\partial t}+\alpha v\cdot\nabla p$$

其中，v 是介质的速度，这种场包含复合的质量传递和扩散．

若流量分布源存在，即系统中含有耗散和某种储蓄参数时，就使扩散方程右边项中加进正比于分布源的一项、对于非均匀场，则有

$$\nabla\cdot\left(\frac{1}{D}\nabla p\right)=E_p\frac{\partial p}{\partial t}-f_i$$

由扩散方程描述的场具有如下特性：

1° 扩散方程所表示的系统兼有耗散参数和某种储蓄参数（只能表现一种储蓄参数）．耗散参数和某种储蓄参数的同时存在，表示出该系统的响应不能立即达到终值或稳态值．常数由系统参数决定，它量度延迟的大小．因此，在这些场中，时间一定是独立变量．

2° 对于包含一种存储元素（而不是两种）的系统，当边界上的激励有陡变时，场内的位势响应将单调地追近终值，即位势梯度的极性没有变化，而且场位势的终值也没有趋调现象．由此，扩散方程有时还称为"调匀方程"，其中追近终值的速率由耗散参数和储蓄参数决定．显然，如果存在稳态条件，也就是说，若激励不是时间的函数，并且当激励变化后，已历经了足够的时间，则扩散方程中的$\frac{\partial p}{\partial t}$为零，由此可以认为拉普拉斯方程是扩散方程的特殊或退化情况．

扩散方程可用来描述含有耗散参数和某一种储蓄参数的场问题．在热传递问题中，通常所研究的系统是含有热阻和热容器的三维场，热阻和热容器分别作为耗散型和储蓄位势参数．尽管热阻不像电阻那样具有能量转换，但从熵的角度来看，可以认为这两

种阻抗在各自范围内起着相似的作用，即使得响应的终值延迟（或阻尼）实现．因此，只要给定边界上的温度（或热流量）以及场内初始时刻的温度分布，就可以由扩散方程确定出场内任意时刻（初始时刻后）各处的温度．

扩散也可以用来描述该空间中某种流体粒子的扩散现象，例如一氧化碳在静止空气中的扩散，墨水在静止水中的扩散．通常，在这些问题中，我们的注意力集中在两种流体中某一种的浓度．扩散也运用了固体吸收液体粒子问题，以及充满液体的多孔固体的干燥问题等．

对于不旋转的，不可压缩的流动问题，由于有粘滞力和惯性力产生，因而也可以用扩散方程预测流动流体中各点的速度位势或压强．不计惯性力的可压缩流体的粘滞流问题，例如多孔介质中的气体流动问题，也可用扩散方程进行分析．

三、双曲型偏微分方程

1. 波动方程

这类方程被称为波动方程，是因为它能描述波的运动．如图6-7所示的一维系统中具有位势和流量储蓄的分布参数，但没有耗散存在．这类系统可以用来表示一种电传输线，其中阻抗不计，但具有显著的对地耦合的电导率和解电率．

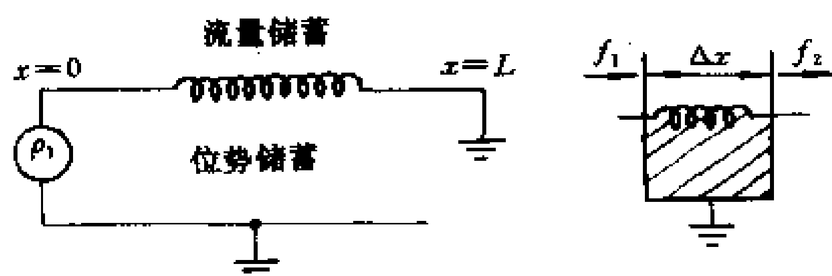


图 6-7　具有流量和位势储蓄的一维场

由位势储蓄的特性，沿系统长度方向的流量变化率为

$$-\frac{\partial f}{\partial x}=E_p\frac{\partial p}{\partial t} \qquad (6)$$

由流量储蓄的特性、位势梯度与流量变化率有关系式

$$-\frac{\partial p}{\partial x}=E_f\frac{\partial f}{\partial t} \tag{7}$$

将(1)式对时间 t 微分,(2)式对 x 微分,得

$$-\frac{\partial^2 f}{\partial x\partial t}=E_p\frac{\partial^2 p}{\partial t^2}$$

$$-\frac{\partial^2 f}{\partial x\partial t}=\frac{1}{E_f}\frac{\partial^2 p}{\partial x^2}$$

联立两方程,有

$$\frac{\partial^2 p}{\partial x^2}=E_pE_f\frac{\partial^2 p}{\partial t^2}$$

这就是一维波动方程. 在求解时,还必须知道边界条件 $p(0,t)=p_0(t)$和 $p(L,t)=p_2(t)$以及两个初始条件,它们相应于 $t=0$ 时刻系统各点的位势和流量,即 $p(x,0)=g(x)$以及位势变化率 $\frac{\partial p(x,0)}{\partial t}=h(x)$.

对于三维情况,有

$$\nabla^2 p=k\frac{\partial^2 p}{\partial t^2}$$

通常,对于不包含耗散或阻尼系数,但是同时具有位势和流量储蓄参数的系统,波动方程总是适用的. 这两储蓄的呈现,意味着物质或能量间的一种内部变化. 加于该系统的能量,或者初始状态时呈现于该系统中的能量,都不会损失掉,其中只存在动能与势能间的来回转变,即形成波动和振荡形式.

在电系统中,没有电阻,但是兼有感抗和容抗的情形,可用波动方程描述,这是理想传输的情况.

在动力学中,对于粘滞阻尼忽略不计,但是兼有惯性力和弹簧力的情况,也可用纯波动方程描述,例如阻尼不计的弦振动和鼓面振动.

在流体力学中,流量存储与各流体粒子的惯性有关,而流体中的位势存储则意味着流体的压缩. 若粘滞力比惯性力小到可以忽

略不计，而流体又是可压缩的，则也可以使用波动方程．

2．阻尼波动方程

上面只讨论了含有两种参数的系统．实际上，许多系统既包括两种储蓄又具有一定幅值的耗散．不失一般性，有必要提供两种耗散：串联耗散 D_s，它反映了流量流引起的能量耗散；并联耗散 D_p，它引入"泄露"造成的系统能量损失．在电气传递中，传输线上有相当大的串联电阻、电感以及两线之间的泄露电导和介电耦合．当用守恒原理去表达 x 坐标方向的流量变化 f_1-f_2 时，必须包括流入位势储蓄器的能量以及贯穿并联耗散元素的能量，即有

$$-\frac{\partial f}{\partial x}=\frac{1}{D_p}p+E_p\frac{\partial p}{\partial t}$$

类似地，位势梯度表示成

$$-\frac{\partial p}{\partial x}=D_sf+E_f\frac{\partial f}{\partial t}$$

上述两方程对 p 的联立解为

$$\frac{\partial^2 p}{\partial x^2}=E_pE_f\frac{\partial^2 p}{\partial t^2}+(D_pE_p+E_fD_s)\frac{\partial p}{\partial t}+D_pD_sp$$

这是大多数一维场的通用表达式，也称为电极方程，因为它描述了沿电波传输线上的电压分布．

显然，若上述方程中 $D_p=D_s=0$，则化为波动方程；若 $E_f=D_p=0$，则为扩散方程；若四种参数 E_f、E_p、D_s 或 D_p 中任意三种为零，则为拉普拉斯方程．方程可一般地写成

$$\nabla^2p=k_1\frac{\partial^2 p}{\partial t^2}+k_2\frac{\partial p}{\partial t}+k_3p$$

它要求有波动方程相同的边界条件和初始条件．

在边界上加入阶跃函数激励，电报方程描述的系统的特征响应是一种阻尼振荡，逐渐地趋近于平衡条件．该过程不一定是单调的（如扩散方程的情形），而是可能在平衡状态的周围产生超调和振荡．从这点上讲，该方程相应于含有电阻、电容和电感的集中参数系统的常微分方程特征，并呈现出欠阻尼、临界阻尼和过阻尼

的特性．

阻尼波动方程可以描述所有呈现出三种参数类型的物理系统．例如，含有较大质量、弹簧力和粘滞阻尼的动力场：一根摆弦或一张松紧面，在受到突然吹动或激励后，形成的逐渐消失的正弦振荡，以及兼有粘滞力和惯性力的可压缩流体等等．

四、双调和方程

双调和方程是弹性理论中的一类特殊方程．在应力分析中，穿透变量（应力）不是矢量，即不能用幅值和方向来定义，它是张量，其中三个分量是 x、y、z 方向的分量，另外三个分量是定义应力参数平面的．实际上只关心两种应力，即法向应力和剪切应力．在弹性力学中，与通用守恒原理相适应的基本定律是平衡方程和相容方程，当使用这些方程去建立弹性体内的应力和变形间的关系时，可以按下列方程定义应力函数 φ：

$$\frac{\partial^2\varphi}{\partial x^2}=\sigma_y,\ \frac{\partial^2\varphi}{\partial y^2}=\sigma_x,\ \frac{\partial^2\varphi}{\partial x\partial y}=\rho_{xy}$$

其中，σ_x 和 σ_y 分别是 x 和 y 方向上的法向应力，ρ_{xy}是相应的剪切应力，在静态条件下，平衡性和相容性就导致所谓双调和方程，其一维形式为

$$\frac{\partial^4\varphi}{\partial x^4}=0$$

二维形式为

$$\frac{\partial^4\varphi}{\partial x^4}+2\frac{\partial^4\varphi}{\partial x^2\partial y^2}+\frac{\partial^4\varphi}{\partial y^4}=0$$

引进双调和算子∇^4，则方程可用紧凑形式表示为

$$\nabla^4\varphi=0$$

在应力问题中，所研究的调和弹性板的重量，相当于内部分布源，当弹性成分的重量较大时，方程修正为

$$\nabla^4\varphi=\omega$$

其中，ω 是单位长度或单位面积的重量．

在瞬态条件下，由杨氏模量表征的弹簧力将起作用，进而呈现振荡，并描述为

$$\nabla^4\varphi=k\frac{\partial^2\varphi}{\partial t^2}$$

在更一般的情况下，方程被修正为

$$\nabla^4\varphi=k_1\frac{\partial^2\varphi}{\partial t^2}+k_2\frac{\partial\varphi}{\partial t}+k_3\varphi$$

§6.4 实 例

例 6-1 地面水源的污染和净化

人们早已认识到，空气和水的污染，给人们精神和健康上带来不良影响是一个严重的问题．70 年代初期，日趋严重的环境问题开始被看作是国内和国际性的危机．这个例子是研究地面水源(湖泊、河流)污染和净化的一个简化模型．

根据 §6.1 中对扩散方程的分析，设水源中污染物的浓度为 $c(x,y,z,t)$，则其变化规律可用下列方程描述：

$$\frac{\partial}{\partial x}(D_x\frac{\partial c}{\partial x})+\frac{\partial}{\partial y}(D_y\frac{\partial c}{\partial y})+\frac{\partial}{\partial z}(D_z\frac{\partial c}{\partial z})$$
$$=\frac{\partial c}{\partial t}+v_x\frac{\partial c}{\partial x}+v_y\frac{\partial c}{\partial y}+v_z\frac{\partial c}{\partial z}+Q(x,y,z,t)+R(x,y,z,t)\quad(1)$$

其中 v_x,v_y,v_z 为水源在 x、y、z 三个方向上的流速；D_x,D_y,D_z 为污染物在 x、y、z 三个方向上的扩散速率(包括分子扩散的随机运动，表现为污染粒子从高浓度点到低浓度点的净运动，以及由于不同的粒子有不同的弯曲路径，造成不规则的流动，从而引起的分散性)；$Q(x,y,z,t)$为污染源；$R(x,y,z,t)$描述了污染物化学衰变．

如果企图计算某一点突然排出污染物的瞬态污染情况，并且不考虑污染物化学衰变．为简化起见，假设源内水沿 x 轴方向流动，即 $v_x=v,v_y=v_z=0,D_x=D_y=D_z=D$. 选取排出污染处为坐

标原点，冲击强度为 Q，则方程可表示为

$$\frac{\partial c}{\partial t}+v\frac{\partial c}{\partial x}-D\left(\frac{\partial^2 c}{\partial x^2}+\frac{\partial^2 c}{\partial y^2}+\frac{\partial^2 c}{\partial z^2}\right)=Q\delta(x,y,z,t)$$

这是扩散方程的初值问题或称自由边界问题.

其解析解为

$$c(x,y,z,t)=\frac{Q}{(4\pi Dt)^{3/2}}\exp\left[-\frac{(x-vt)^2+y^2+z^2}{4Dt}\right]$$

一般的问题可以用数值方法求解.

对于那些较窄的河流，方程可进一步简化为

$$D\frac{\partial c}{\partial x^2}=\frac{\partial c}{\partial t}+v\frac{\partial c}{\partial x}+Q\delta(x,t)$$

利用这一方程，计算一段时间内大块污染物顺流而下的扩散情况如图 6-8 所示.

对于采用注入清水净化湖泊问题，可对(1)式求数值解，其中 $Q=Q(t)$，表示在坐标原点注入清水，即可得到在任意时刻 t，任何空间位置点处的污染物浓度.从而判断净化效果.

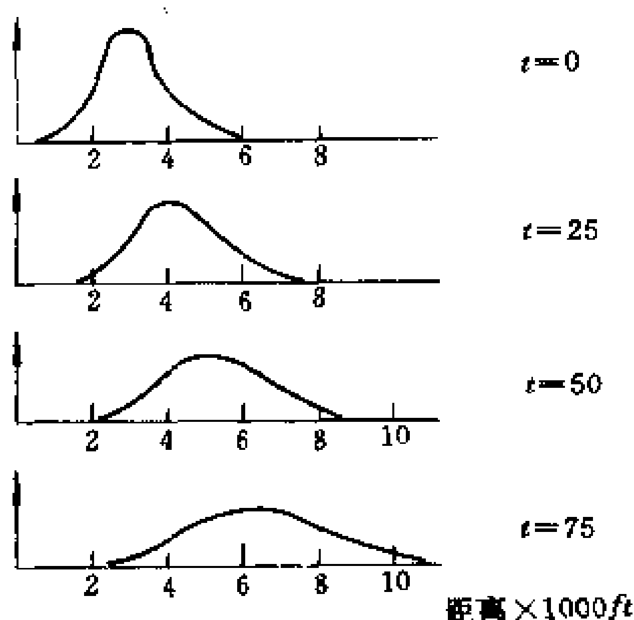


图 6-8　污染物顺流而下扩散的浓度分布图

若建模的目的，主要是为了提高决策能力的话，则从定性和定量测量的观点来看，并不认为偏微分方程模型是必不可少的．在净化问题中，如果只是希望粗略地估计注水时，污染物浓度的变化规律，并决定注水时间，可以采用集中参数的方法来近似处理这一分布参数系统．

假设：1° 水源的蒸发量与降雨量相等，流入与流出水源的平均速率相等，因而水源总量保持不变；2° 污水不发生生化反应，也不引起沉积；3° 污水在水源中瞬时混合均匀，即水源中污水的浓度总是均匀的．

设 t 时刻水源中污水浓度为 $c(t)$，注入水的速率 v 及注入水中污水浓度 c_i 均为常量，记水源体积为 V，水源中污水初始浓度为 c_0.

考虑在 Δt 时间段内，由于污染物的平衡，可得

$$[c(t+\Delta t)-c(t)]V=[c_i-c(t)]V\Delta t+0(V)$$

当 $\Delta t\to 0$ 时，有

$$\begin{cases} V\dfrac{\mathrm{d}c}{\mathrm{d}t}=v(c_i-c) \\ c(0)=c_0 \end{cases}$$

这是一阶线性常系数微分方程，其解为

$$c(t)=c_i-(c_i-c_0)\mathrm{e}^{-t/\tau}$$

且 $\tau=V/v$.

若以清水注入水源，问需多长时间才能使水源中污水浓度降到初始浓度的10%？即令 $c_i=0, c(t)=0.1c_0$，则

$$0.1c_0=c_0\mathrm{e}^{-t/\tau} \quad t=-\tau\cdot\ln 0.1\approx 2.3V/v.$$

例 6-2 在冻土地区施工

在冻土地区施工面临着一个严峻的问题：如果建筑物下面的冰融化了，那么建筑物将陷入土壤中去．根据国外在北极施工数十年的经验，解决方法是使建筑物下面的土壤顶层永久冰冻．具体措施是用某种方式使建筑物和大地隔热，以防止因冰的融化所

引起的支承强度的降低，因此建筑物的地基由沙、石子、木料和其他人造材料装配到不同厚度的各个夹层中构成．在结构设计时，需要一套方法来判别给出的设计能否符合要求，即防止松散的富含水分的土壤解冻．判别方法的主要组成部分是测定地面温度状况，并据此预测地下温度状况．

一、基本模型的建立

根据问题的提法和§6.1的分析，设$c(x,y,z,t)$为地下(x,y,z)处t时刻的温度，则可描述为

$$\frac{\partial c}{\partial t}=\frac{\partial}{\partial x}\left(D_x\frac{\partial c}{\partial x}\right)+\frac{\partial}{\partial y}\left(D_y\frac{\partial c}{\partial y}\right)+\frac{\partial}{\partial z}\left(D_z\frac{\partial c}{\partial z}\right).$$

$$c(x,y,z,t)\big|_{x=0}=c_0(y,z,t),c(x,y,z,0)=f(x,y,z)$$

其中坐标系的选取如图6-9所示．

这一模型在处理时计算量很大，考虑将模型简化．

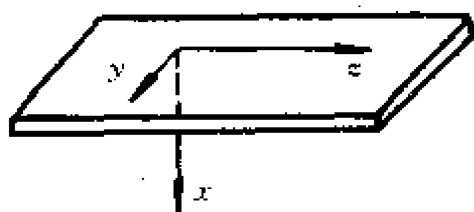


图6-9　坐标系的选取

1°忽略边缘效应．根据多年的经验，建筑物的边缘只在4～5ft的范围内受到边缘热效应的作用，而建筑物在每个方向上可以延伸30ft或者更多，所以在建筑物主要部分下的湿度场只与深度有关．这时温度场的模型只是时间和浓度的函数．

2°只考虑热传导．根据沙子和砾石的特性，三种形式的热传递（辐射、对流和传导）确实都呈现在渗水的泥土中．不过经验证明，在这样一种介质中的热传递，都可以用单纯热传导的模型作充分的描述．

3°层状假设．设建筑物的地基由很多层厚度不一的厚块组成，每层底部为水平，保持在同一温度上．每一层内介质为均匀的．

考虑没有水分的由同类材料构成的单层模型：

$$k\frac{\partial^2 u(x,t)}{\partial x^2}=c\rho\frac{\partial u(x,t)}{\partial t}$$

其中，c 为该层介质的比热，ρ 为密度，k 为传导率.

在观察开始时的地下初始温度分布

$$u(x,0)=u_0(x)$$

边值条件:

地表温度　$u(0,t)=a(t)$

介质底层温度　$u(X,t)=u_X$　(在底部 $x=X$ 处，温度将固定在某一温度 u_X，例如 $u_X=32°F$.

实际上，地表温度具有随机性，此处为简单起见，认为地表温度是时间的一个确定函数.

以上模型是一个适定的初边值混合问题. 要求解这个模型，必须由观测和试验获得函数 $a(t)$，$u_0(t)$以及物理常数 k,c,ρ.

二、单层模型的深化

现考虑单层的均匀模型，但含有冰水混杂状态的水分，如图 6-10 所示.

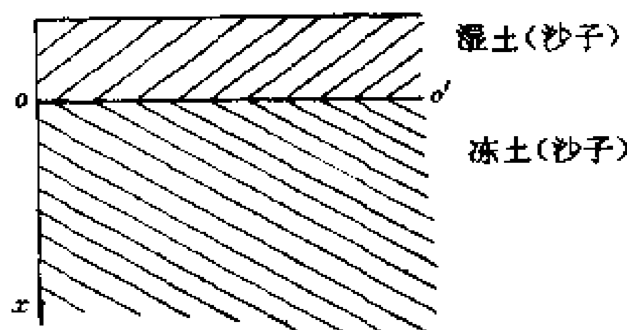


图 6-10　单层模型的深化

设在已知时刻 t_0，沙子在直线 00' 以上是潮湿的，而在 00' 以下是冻结的. 显然，这要求 $a(t_0)>32°F$，$u_X<32°F$. 将沙子看作两层，对 00' 以上部分，有

$$k_1\frac{\partial^2 u}{\partial x^2}=c_1\rho_1\frac{\partial u}{\partial t},u(0,t)=a(t)$$

其中 k_1, c_1, ρ_1 的值依据潮湿沙子测定，而对直线 OO^1 以下部分，有

$$k_2 \frac{\partial^2 u}{\partial x^2} = c_2 \rho_2 \frac{\partial u}{\partial t}, u(X, t) = u_x,$$

其中 k_2, c_2, ρ_2 的值依据冻结沙子测定．

为了计算未来时刻的临界分布，必须耦合自由界面 OO^1 处潮湿沙子和冻结沙子间的温度和热通量．

温度情况是简单的，潮湿沙子层底部和冻结沙子层顶部都为水的冻结温度(32°F)．所以

$$u_{sw}(s(t), t) = u_{sI}(s(t), t) = 32°\text{F}$$

其中，$s(t)$表示自由界面 OO^1，sw 为潮湿沙子(Sand wet)，s_I 为冻结沙子(Sand ice.)．这是第一个自由边界条件．

自然，自由界面 $s(t)$随着时间推移会移动．假定在时间间隔 Δt 内，界面 $s(t)$向下运动了距离 Δs，那么隔化一块冻结沙子 $\Delta s \cdot A$(其中 A 为地基面积)用去热量为

$$\Delta Q = \rho_{sI} \lambda \Delta s \cdot A$$

其中 ρ_{SI}为冻结沙子的密度，λ 为冻结沙子的融解热．

而融化所需热量，由从潮湿沙子流出的热量减去传导进冰冻沙子的热损失提供，故有

$$A k_{SI} \frac{\partial u_{SI}}{\partial x} - A k_{Sw} \frac{\partial u_{Sw}}{\partial x} = \rho_{SI} \lambda \frac{\Delta s}{\Delta t} \cdot A$$

这里热通量的计算，利用了热传导的 Fourier 法则：通过一个面的热流率与这个面的面积和垂直于这个面的温度梯度成正比，即 $\frac{\Delta Q}{\Delta t} = -kA \frac{\partial u}{\partial x}$.

令 $\Delta t \to 0$，得到了第二个自由边界条件：

$$k_{SI} \frac{\partial u_{SI}}{\partial x} - k_{SW} \frac{\partial u_{SW}}{\partial x} = \rho_{SI} \lambda s'(t)$$

为了找到部分冻结的均匀土质中的温度分布，需要同时求解 $s(t)$上面和下面的两个热传导方程服从已知边界和自由界面条件

的解$\{u_{SW}, u_{SI}, s(t)\}$.

三、多层模型

设研究的结构由均匀的但不相同的若干层组成，这些层可以依次建立在层状土壤上．对每一个均匀层，上面的单层模型都适用．在两层之间的边界上，假定为理想连接，即温度和热通量是连续的．例如，如果在深度$x=x_k$处，有一个沙子层和石子层的界面，则

$$\lim_{\varepsilon\to 0}[u(x_k+\varepsilon,t)-u(x_k-\varepsilon,t)]=0$$

$$\lim_{\varepsilon\to 0}\left[k_g\frac{\partial u}{\partial x}(x_k+\varepsilon,t)-k_s\frac{\partial u}{\partial x}(x_k-\varepsilon,t)\right]=0$$

其中，下标s代表沙子(sand)，下标g代表砂砾(grit)，传导率k_g和k_s的值与沙子和砾石在x_k处是否为潮湿或冻结有关．

综上所述，层状介质中热传导的数学模型可以描述如下．

假设共有m个固定的不相同的层，对从深度x_{i-1}到x_i的第i层，热量参数定义如下：

k_i^f或k_i^w：冻结(frozen)或潮湿(wef)时第i层的热导率．

ρ_i：　第i层的密度，假定对于潮湿和冻结的土壤相同．

c_i^f或c_i^w：冻结或潮湿的第i层热容量．

λ_i：　第i层的融解热．

记$u(x,t)$为在x处t时刻的温度；$s(t)$为冻土和湿土之间的界面；$a(t)$为地表($x=0$)处的温度；u_X为最后一层底部($x=x_m$)处的不变温度($u_X<32°F$)；$u_0(x)$为开始计算时各层的温度分布，则层状介质中的温度分布由下述自由边界问题描述：

$$k_i^r\frac{\partial^2 u}{\partial x^2}-\rho_i^r\rho_i\frac{\partial u}{\partial t}=0$$

其中，
$$r=\begin{cases}w, x\in(0,s(t))\\ f, x\in(s(t),X)\end{cases}$$

表面边界条件　$u(0,t)=a(t)$

在固定界面上的热流条件

$$\lim_{\varepsilon\to 0}[u(x_i-\varepsilon,t)-u(x_i+\varepsilon,t)]=0$$

$$\lim_{\varepsilon\to 0}[k_i^r\frac{\partial u}{\partial x}(x_i-\varepsilon,t)-k_{i+1}^r\frac{\partial u}{\partial x}(x_i+\varepsilon,t)]=0$$

底部边界条件　　$u(X,t)=u_X$

自由边界条件　　$\lim_{\varepsilon\to 0}u(s(t)+\varepsilon,t)=\lim_{\varepsilon\to 0}u(s(t)-\varepsilon,t)=0$

$$\lim_{\varepsilon\to 0}[k_j^f\frac{\partial u}{\partial x}(s(t)+\varepsilon,t)-k_j^w\frac{\partial u}{\partial x}(s(t)-\varepsilon,t)]=\rho_j\lambda_j\frac{\mathrm{d}s}{\mathrm{d}t}$$

（其中 j 为含有 $s(t)$层的标号）

初值条件　　$u(x,0)=u_0(x)$

$u_0(x)=0$ 处的是 s_0 是自由界面的初始位置，因此 $s(0)=s_0$.

这一问题称为双相斯忒藩(stefan)问题，以纪念奥地利物理学家 J. Stefan. 他用这种模型研究了冰——水系统. 这种类型的问题在化学反映、生物的扩散、粘弹性的扩散和恒星演化中得到应用.

下面是对永久冻土上人造的四层结构中，对地下温度的预测，如图 6-11 所示：

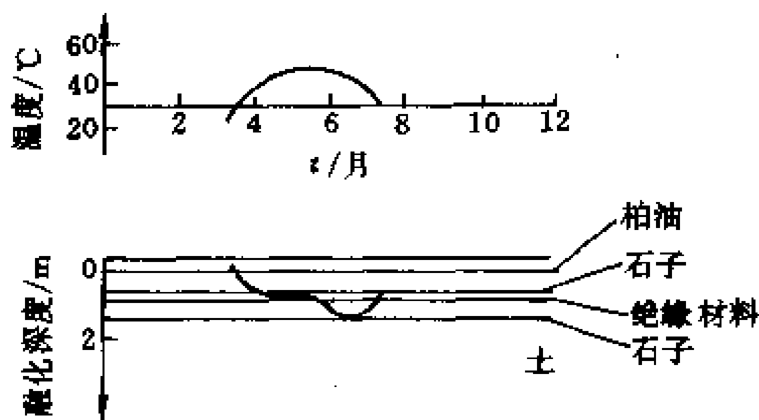


图 6-11　地表温度变化时融化深度图示

例 6-3　航道问题

现测得一矩形水域(75200)×(−50,150)中的某些点水深如

表(单位:米),如果你的船吃水为 5 米,问你应避免进入哪些区域?

表 6-2 航道水深测量数据

x	129.0	140.0	108.5	88.0	185.5	195.0	105.5	157.5
y	7.5	141.5	28.0	147.0	22.5	137.5	85.5	−6.5
z	4	8	6	8	6	8	8	9

x	107.5	77.0	81.0	162.0	162.0	117.5
y	−81.0	3.0	56.5	−66.5	84.0	−38.5
z	9	8	8	9	4	9

实际上,只要能画出该水域的等深线图或三维地形图,问题即可解决. 因此这是一个根据部分测量数据(x_k,y_k,f_k),$k=1,2,\cdots,n$,作曲面拟合的问题. 而利用三次样条作曲面拟合,必须知道网格节点上的值.

已知部分网格节点上的值$f_k(x_k,y_k)$,$k=1,2,\cdots,n$,求其他网格节点上的值,这是插值问题. 处理平面插值问题有三种主要方法:

1. 加权法

任意点(x,y)处的函数值$F(x,y)$,可利用周围已知点(x_k,y_k)的函数值f_k加权得到

$$F(x,y)=\sum_{k=1}^{n}w_k(x,y)f_k\Big/\sum_{k=1}^{n}w_k(x,y)$$

其中,$w_k=1/[(x-x_k)^2+(y-y_k)^2]$.

2. 最小二乘法

选取适当的模型函数$\tilde{F}(a_0,a_1,\cdots,a_n,x,y)$,使

$$\min\sum_{k=1}^{n}[f_k-\tilde{F}(a_0,a_1,\cdots,a_n,x_k,y_k)]^2w_k(x_k,y_k)$$

其中,$\tilde{F}$可选为三次样条.

3. 直接解差分方程

将曲面认为是某偏微分方程的解曲面，求解此偏微分方程，可得网格节点上的值．

具体分析该问题所提供的测量数据，在网格上是稀疏的，如图 6-12 所示．

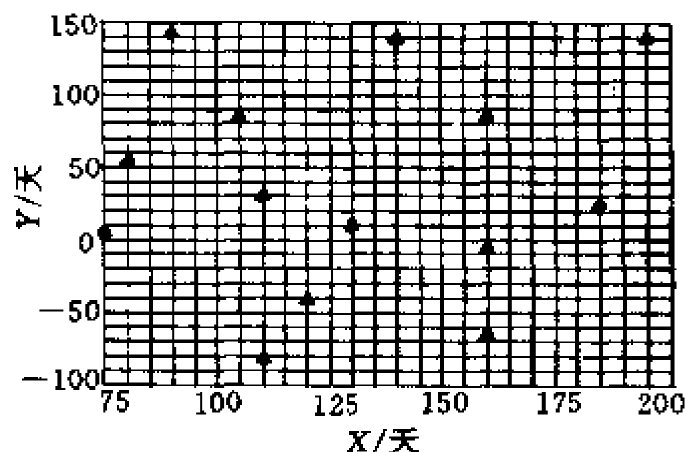


图 6-12　测量数据在网格上的图示

考虑采用第三种方法．将水域底面设想为一矩形板，在 $(x_k, y_k)k=1,2,\cdots,n$ 处受到力的作用，产生形变，设 $u(x,y)$ 为位移，则满足 § 6.3 中分析的双调和方程

$$\frac{\partial^4 u}{\partial x^4}+2\frac{\partial^4 u}{\partial x^2\partial y^2}+\frac{\partial^4 u}{\partial y^4}=\begin{cases} f_k & x=x_k, y=y_k, \quad k=1,2,\cdots,n \\ 0 & \text{其他} \end{cases}$$

自由边界条件，

采用中心差分将 $\frac{\partial^4 u}{\partial x^4}, \frac{\partial^4 u}{\partial y^4}, \frac{\partial^4 u}{\partial x^2\partial y^2}$ 离散化，得有限差分方程：

1° 对于域内部网格节点，有

$$u_{i+2,j}+u_{i,j+2}+u_{i-2,j}+u_{i,j-2}+2(u_{i-1,j+1}+u_{i+1,j+1}+u_{i+1,j-1}$$
$$+u_{i-1,j-1})-8(u_{i+1,j}+u_{i-1,j}+u_{i,j+1}+u_{i,j-1})+20u_{ij}=0$$

2° 对于边界上的节点，

$$u_{i-2,j}+u_{i+2,j}+u_{i,j+2}+u_{i-1,j+1}+u_{i+1,j+1}$$
$$-4(u_{i-1,j}+u_{i,j+1}+4u_{i-1,j})+7u_{i,j}=0$$

3° 对于有一列来自某一边界的节点，

$$u_{i-2,j}+u_{i+2,j}+u_{i,j+2}+2(u_{i-1,j+1}+u_{i+1,j+1})+(u_{i-1,j-1}$$

$$+u_{i+1,j-1})-8(u_{i-1,j}+u_{i,j+1}+u_{i+1,j})-4u_{i,j+1}+19u_{ij}=0$$

4° 对于角落上的节点，

$$2u_{i,j}+u_{i,j+2}+u_{i-2,j}-2(u_{i,j+1}+u_{i+1,j})=0$$

5° 对于紧挨着一个角落的节点

$$u_{i,j+2}+u_{i-2,j}+u_{i-1,j+1}+u_{i+1,j-1}+2u_{i-1,j-1}$$
$$-8(u_{i,j+1}+u_{i+1,j})-4(u_{i,j-1}+u_{i-1,j})+18u_{i,j}=0$$

6° 对于在一边上紧挨着一个角落的节点，

$$u_{i,j-2}+u_{i+1,j-1}+u_{i-1,j+1}+u_{i-2,j}-2u_{i-1,j}$$
$$-4(u_{i+1,j}+u_{i,j+1})+6u_{i,j}=0$$

迭代后，获得网格节点的值，即偏微分方程的数值解．利用三次样条作曲面拟合，得水域地形图，问题得到解决，如图 6-13 所示．

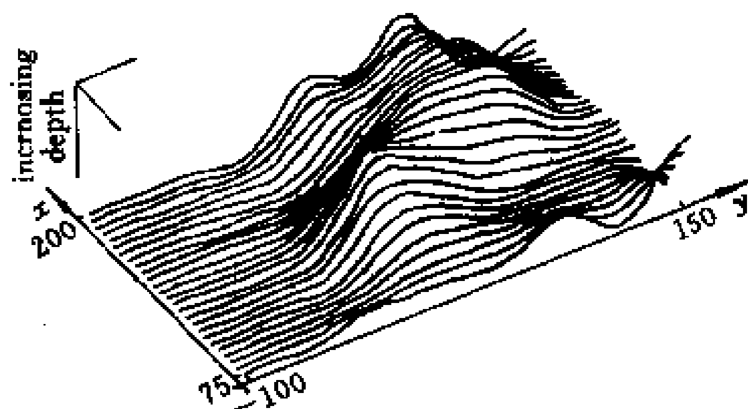


图 6-13　河床地形图

§ 6.5　数学物理反问题

在各种不同的学科领域中，经常会遇到量之间的转化以及问题的"对称"．最常见的量的转化是已知量与未知量的转化，例如原问题的已知量变成了新问题的未知量，可以把这样一对问题称为正、反问题：例如在代数中如果将已知方程求根称为正问题的话，那么由根求方程的系数就是代数方程的反问题．

微分方程是描述与刻划物理过程、系统状态、社会与生物现象的有力工具. 牛顿(Newton)说“探索自然界的奥秘在于解微分方程”. 这种由“原因”推得“结果”无疑在人类认识自然与改造自然中起到了重要作用. 微分方程反问题是指由“果”反推“因”,即已知或部分已知微分方程的解反求方程中的未知部分.

在前面已经看到,偏微分方程的定解问题——正问题,通常是寻求满足支配方程、初始条件与边界条件(对某些方程可能没有初始条件或边界条件)的解. 如果在微分方程定解问题中原来已知量中的某一个或某几个变成了未知量,而原来未知的函数现在或者是已知的,或者是与这未知函数有关的某些信息是已知的,称这些信息为“附加条件”. 微分方程反问题是指由支配方程,初边值条件和附加条件去确定原问题中的未知量. 附加条件的给法在微分方程反问题中十分重要,在理论上应保证解的存在性和唯一性(有时不要求解的唯一性),在实践中则要求这些量的可观测性与易观测性.

从工程角度,微分方程反问题可分为“控制”、“识别”、“综合”等类型,在数学上则可大致上分为如下四类.

第一类　算子识别

这一类指的是待定微分方程中的含有未知参数的反问题,在自然科学与工程技术的各个领域中最为常见. 在工程中,由于直接测量这些参数在离散点处的值往往不太容易,不得不转而去测量与待定参数有一定关系的其他量在边界上的值或其他可获得的信息,去推断待求的量. 工程中常称这类问题为遥感.

第二类　逆时间过程问题

这是指一类待定初始条件的反问题,例如在扩散方程中,由函数在 t_0 时刻的值去求 $t=0$ 时刻的值,这种问题常常是非适定的.

第三类　边界控制问题

这是一类待定未知边界条件的反问题,即为了使微分方程的解具有某些性质应当如何设置边界(或部分边界条件)条件的问

题，这类问题也常常是不适定的．

第四类　几何反问题

这类问题中微分方程定解问题中的边界形状是待求量，工程中的某些定向设计问题常可归为这类反问题，数学物理中遇到的确定不同界质交界面的形状也可归为这类反问题．这类问题还包括工程中常常遇到的“杂交问题”，即边界的某一部分是未知的，在其上规定某种边界条件，而边界的其余部分是已知的，并像通常的正问题一样给定边界条件（亦称为“混合问题”）．

下面以一个简单的例子说明数学物理反问题的大致处理过程．

例 6-4　脑病研究中需要测试新药的疗效，例如治疗帕金森症的多巴胺的脑部注射疗效．为了精确估计药物影响的脑部范围，必须估计注射后药物空间分布的形状和大小．

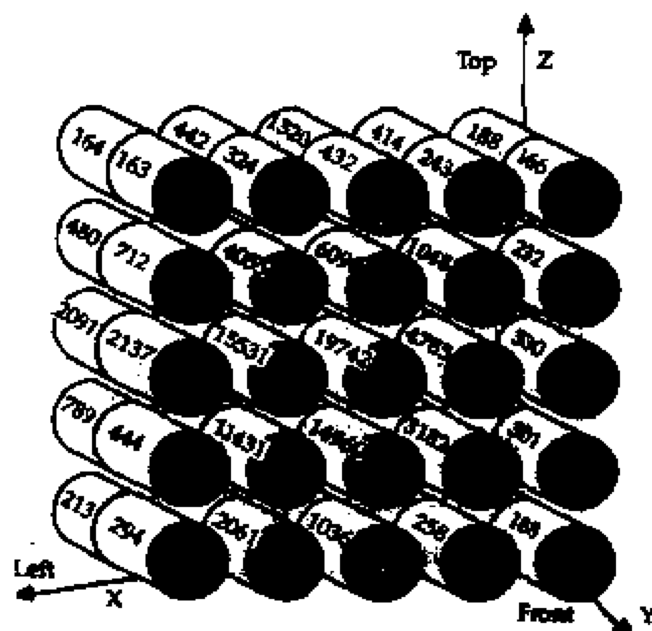


图 6-14　由耗散物质构成的长方体

试验数据包括 50 个圆柱形组织样本中每个药物含量的测量值（见图 6-14 和表 6-3）．每个圆柱体高 0.76mm，直径为 0.9mm．这些平行圆柱的中心位于 1.00mm×0.76mm×1.00mm 的格点上，各圆柱体的底面彼此相连，而侧面互不接触．注射点位于药剂

量最高的圆柱中心附近，当然在圆柱体组织之间及样本范围之外的地方也有药物．试验数据用计数单位表示(1 单位含 4.735×10^{-13}克分子药量)．

表 6-3　圆柱形组织样本中药物含量测量值

后方垂直截面					前方垂直截面				
164	442	1320	414	188	163	324	432	243	166
480	7022	14411	5158	352	712	4055	6098	1048	232
2091	23027	28353	13138	681	2137	15531	19742	4785	330
789	21260	20921	11731	727	444	11431	14960	3182	301
213	1303	3765	1715	453	294	2061	1036	258	188

记 $c=c(x,y,z,t)$ 是 t 时刻点 (x,y,z) 处的多巴胺的浓度．假设

1° 大脑或神经系统中的多巴胺起着神经传输的功能，数量不等地存在于整个大脑中，它可以彼综合也可以被代谢．例如，老鼠脑中的多巴胺的正常浓度是每 mm^3 脑组织中含 10—119 计数单位．基于这一事实，假设注射前大脑中多巴胺的含量可以忽略不计．

2° 多巴胺在大脑中的分布由扩散和衰减过程所确定，而忽略其他过程，并认为大脑是均匀的，沿 x,y,z 轴方向的扩散系数分别为常数，还认为这一扩散过程服从 Nernst 实验定律：在 Δt 时间内流过曲面 Δs 的粒子数与浓度沿 s 的法向导数成正比．衰减过程中，质量的减少与浓度成正比．

3° 假定注射点在后排中央那个圆柱体的中心，且这个中心位于 $y=0$ 平面上．取样时间是未知的，完成取样与注射的时间长短是可以忽略的，即看成瞬时完成的．

4° 多巴胺的追踪试验通常是在实验室中对动物进行的，取样部分只占很小一部分，而且不取在大脑边界附近，因而不考虑大脑

边界面的影响．

根据上述假设，可建立一阶衰减的非稳定态的质量扩散运输方程

$$\frac{\partial c}{\partial t}=\frac{\partial}{\partial x}\left(E_x\frac{\partial c}{\partial x}\right)+\frac{\partial}{\partial y}\left(E_y\frac{\partial c}{\partial y}\right)+\frac{\partial}{\partial z}\left(E_z\frac{\partial c}{\partial z}\right)-kc$$
$$+\delta M(x-x_0)(y-y_0)(z-z_0)$$

其中，E_x,E_y,E_z 是 x,y,z 方向相应的扩散系数，k 是衰减系数，M 为瞬时点源的质量 x_0,y_0,z_0 为点源坐标．上述方程的解析解为

$$c(x,y,z,t)=\frac{M}{(4\pi t)^{3/2}(E_xE_yE_z)^{1/2}}$$
$$\cdot\exp\left\{-\frac{(x-x_0)^2}{4tE_x}-\frac{(y-y_0)^2}{4tE_y}-\frac{(z-z_0)^2}{4tE_z}-Kt\right\}$$

1. 数值求积方案

因为数据是用圆柱取样上的计数单位表示的，而模型预测了多巴胺的浓度值，有必要取应变量 c 在每个圆柱体上的积分值，使之能与给定数据相比较. 由于模型的解对空间变量的积分没有解析表达式，故采用数值积分．记第 i 个柱体上的多巴胺克分子质量为 m_i，则

$$m_i=\iiint\limits_{v_i}c(x,y,z)\mathrm{d}v_i\approx\sum_j c_{ij}\Delta x\Delta y\Delta z$$

其中 c_{ij}是第 i 个圆柱体的第 j 个“微元”上多巴胺浓度的计算值．令样本柱体的直径为 D，取 $\Delta x=\Delta y=\Delta z=D/6$.

2. 参数估计

模型中有 8 个参数：$x_0,y_0,z_0,M,E_x,E_y,E_z,k$.

取 $y_0=L/2$，即假定注射是在与样本的后排垂直截面的二分面上进行的．

由于 x_0,z_0 的值不是用模型和数据相比较而得到的，所以作特殊处理．令

$$c(x,z)=ax^2+bz^2+cxz+dx+ez+f$$

其中，c 表示在 x,z 处多巴胺的回归值，a,b,c,d,e,f 均为常数值，

求使 $c(x,z)$ 达到最大值的 x,z 坐标 x_0,z_0，即令 $\frac{\partial c}{\partial x}=0,\frac{\partial c}{\partial z}=0$，得到 x_0 和 z_0 的估计值.

表 6-4

	x_0	z_0
后排垂直截面	3.271	2.686
前排垂直截面	3.265	2.766
平均值	3.628	2.726

由假设，实际注射量 M 为已知(若作未知参数，则 M 与 k 的强烈正相关，使处理变得非常复杂). 由实测数据，在格点上搜索，估计 M 的范围在 1×10^6 到 1.5×10^6 多巴胺计数，因此假定注射多巴胺的 M 值为 10mg(约为 1.4×10^6 计数单位).

据网络搜索发现，E_x 与 E_z 高度相关，因此假设 $E_x=E_z$，问题转化为一个三参数的优化问题.

选取 E_x,E_y,k，使 $\min\sum_i(\hat{m}_i-m_i)^2$，解得 $E_x=E_z=0.375,E_y=0.275,k=0.140$.

3. 结果比较

将表 6-3 与表 6-5 比较，可看到预测误差.

表 6-5 多巴胺预测值(计数)

后部垂直截面					前部垂直截面				
153	753	1087	406	42	95	466	672	251	26
1518	7454	10808	4042	419	939	4609	6682	2499	259
4424	21796	31534	11764	1213	2735	13476	19497	7273	750
3312	16320	23593	8800	907	2048	10090	14587	5441	561
688	3394	4886	1821	187	426	2098	3021	1126	116

该问题的统计分析可参考文献[2].

从数学角度，微分方程反问题的研究大体包含以下三个方面. 第一，在理论上解决反问题提法的正确性问题. 通常反问题的完整描述应该包括微分方程、定解条件（它们组成微分方程的正问题）以及附加的条件，附加条件一般是指对解的限制以及待定参量的边界条件等. 附加条件的给定要能保证反问题的解是存在唯一的. 第二，反问题的求解方法，对于微分方程的正问题已经有一系列的解析或数值方法，但这些方法通常是不能直接用来求解反问题的，因此对于不同类型的反问题必须通过适当的运算与变换化成可计算的模式，例如积分方程、微分方程或级数等形式. 由于反问题的非适定性，这种经转换得到的计算模式通常也是非适定的. 第三，解决非适定问题的计算稳定性问题. 这些问题的深入研究可参考文献［38］［39］.

习　题

1. 试分别说明弦振动方程$\frac{\partial^2 u}{\partial t^2}-a^2\frac{\partial^2 u}{\partial x^2}=0(0<x<l,t>0)$和热传导方程$\frac{\partial u}{\partial t}-a^2\frac{\partial^2 u}{\partial x^2}=0, x\in(-\infty,+\infty), t>0$中参数$a$的物理意义.

2. 由两根不同质料的杆接成一根杆，当杆作纵振动时，推导在连结点c处应满足的连接条件.

3. 对弦振动（纵振动）问题，分别对（1）两端固定；（2）一端固定，而另一端受一个沿x方向的已知外力作用，以及（3）一端固定，另一端受到一弹簧施加的力的作用（弹簧的一端固定、另一端连接在弦的自由端上），建立边界条件.

4. 图 6-15 是一个由具有均匀特性的耗散性物质构成的长方体，在该长方体左、右、顶三个侧面上均加有不同常数值的温度场. 试建立数学模型. 若三个侧面上的温度场为随时间变化的函数，数学模型有什么变化.

5. 混凝土浇灌后逐渐放出“水化热”，放热速率正比于当时尚储存着的水化热密度，即$\frac{dQ}{dt}=-\beta Q$，试推导浇灌后的混凝土内的热传导方程.

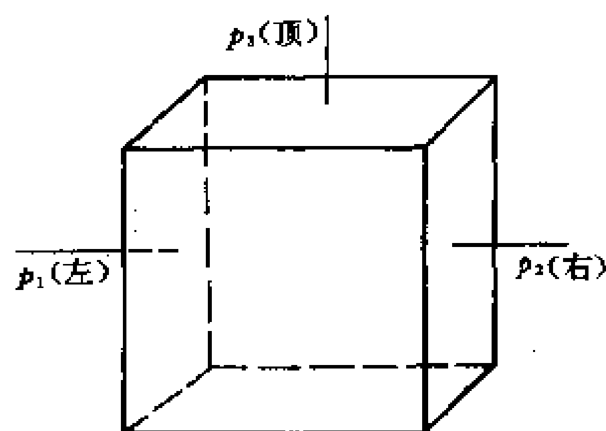


图 6-15　由耗散物质构成的长方体

6. 在杆的纵向振动中，如杆的密度 ρ 和杨氏模量 E 都不是均匀的，而是 x 的连续函数，杆的纵振动方程有何改变？

7. 比较热传导方程的推导和杆的纵振动方程的推导，说明其异同．

第二部分 数据分析法

第七章 回归分析法

回归分析法是"黑箱"建模中常用的方法，用于对函数 $f(x)$ 的一组观测值 (x_i, f_i)，$i=1,2,\cdots,m$. 确定函数的表示．当然 $f(x)$ 不仅是一元的，也可以是多元的．

§7.1 一元线性回归

一元线性回归是处理两个变量之间关系的最简单的模型．它虽然比较简单，却能反映回归分析方法的基本思想和应用．

对一组观测数据 (x_i, y_i)，$i=1,2,\cdots,m$，对于两变量 x 和 y 之间的关系事先一无所知的情况下，通常是首先画出散点图，以便估计其中的关系．

例 7-1 为了估计山上积雪融化后对下游灌溉的影响，在山上建立了一个观察站，测量了最大积雪深度 (x) 与当年灌溉面积 (y)，得到连续 10 年的数据如表 7-1 所示．

表 7-1 最大积雪深度与灌溉面积观测数据

年序	最大积雪深度 x(尺)	灌溉面积 y(千亩)
1	15.2	28.6
2	10.4	19.3
3	21.2	40.5
4	18.6	35.6

续表

年序	最大积雪深度 x(尺)	灌溉面积 y(千亩)
5	26.4	48.9
6	23.4	45.0
7	13.5	29.2
8	16.7	34.1
9	24.0	46.7
10	19.1	37.4

为了研究这些数据中所蕴含的规律性，把各年最大积雪深度作横坐标，相应的灌溉面积作纵坐标，将这些数据点标在平面直角坐标图上．如图 7-1 所示．

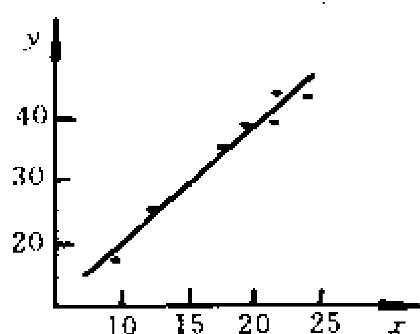


图 7-1

从图中可以看到，数据点大致落在一条直线附近，因此假设有如下结构式

$$y=\alpha+\beta x+\varepsilon$$

其中，α 和 β 是未知常数，称为回归系数，ε 是随机误差，满足 $E(\varepsilon)=0, \mathrm{var}(\varepsilon)=\sigma^2$．只要能从观测数据中得到 α 和 β 的估计值 a 和 b，则回归方程为

$$\hat{y}=a+bx$$

它近似地反映了变量 x 和 y 之间的线性关系，其中 a 和 b 亦称为回归系数．

选择 a 和 b 的原则是使回归直线与所有数据点都比较"接近"，通常采用残差平方和来刻划所有观察值与回归直线的偏差程度，即选择 a 和 b，使得

$$\min Q=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-a-bx_i)^2$$

根据微积分中求极值的方法，应有

$$\begin{cases}\dfrac{\partial Q}{\partial a}=-2\sum_{i=1}^{n}(y_i-a-bx_i)=0\\ \dfrac{\partial Q}{\partial b}=-2\sum_{i=1}^{n}(y_i-a-bx_i)x_i=0\end{cases}$$

得

$$a=\overline{y}-b\overline{x}\qquad b=\frac{L_{xy}}{L_{xx}}$$

其中，

$$\overline{x}=\frac{1}{h}\sum_{i=1}^{h}x_i,\overline{y}=\frac{1}{h}\sum_{i=1}^{h}y_i$$

$$L_{xx}=\sum_{i=1}^{h}(x_i-\overline{x})^2=\sum_{i=1}^{h}x_i^2-\frac{1}{h}\Big(\sum_{i=1}^{h}x_i\Big)^2$$

$$L_{xy}=\sum_{i=1}^{h}(x_i-\overline{x})(y_i-\overline{y})=\sum_{i=1}^{h}x_iy_i-\frac{1}{h}\Big(\sum_{i=1}^{h}x_i\Big)\Big(\sum_{i=1}^{h}y_i\Big)$$

利用上述公式求得例 7-1 中，$b=1.813$，$a=2.355$，回归方程为

$$\hat{y}=2.355+1.813x$$

从图 7-1 上可以看出这个回归方程的图像，与所有数据点都很接近．

一元线性回归模型，应用极为广泛．但是这一方法有局限性，其局限性反映在假设条件中：1° $E(\epsilon)=0$，2° $\mathrm{var}(\epsilon)=\sigma^2$，3° x 与 y 是线性关系．因此当上述条件不满足时，处理方法应作适当修改．

若 $E(\epsilon)=c\neq0$，可作平移处理．当该模型推广到多元情况，若 $\mathrm{cov}(\epsilon_i,\epsilon_j)=\sigma_n^2I$ 不满足，而是 $\mathrm{cov}(\epsilon_i,\epsilon_j)=\Omega$，$\Omega$ 为一非奇异对称阵，则采用广义最小二乘法(GLS)．在多元情况，若 $x_1,x_2,\cdots,x_k$，相关，即对多重共线性问题，应采用岭回归及其变种方法处理．而线性假设，来源于对散点图的观察和对模型精度的要求．两变量

之间是否有线性关系，可采用 F 检验，即如果 x 与 y 有线性关系，则

$$F=\frac{S_{回}/1}{S_{残}/(n-2)}\sim F(1,n-2)$$

其中，$S_{回}=\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2$ 称为回归平方和，$S_{残}=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ 称为残差平方和，$F(1,n-2)$表示第一自由度为 1，第二自由度为 $n-2$ 的 F 分布．这时，在给定显著性水平 α 下，F 值应大于 F 表中的临界值 F_{α}．对例 7-1 进行检验，

$$F=\frac{748961}{16.074/(10-2)}=372.766>11.26=F_{0.01}(1,8)$$

这表明 x 与 y 之间有十分显著的线性相关关系．两变量是否具有线性关系的另一检验方法是利用相关系数 $r=\frac{L_{xy}}{\sqrt{L_{xx}}\sqrt{L_{yy}}}$，查阅相关系数检验表．通常当$|r|$大于表上 $\alpha=5\%$相应值，但小于表上 $\alpha=1\%$相应值时，称 x 与 y 有显著的线性关系；如果$|r|$大于表上 $\alpha=1\%$相应的值时，称 x 与 y 有十分显著的线性关系；如果$|r|$小于表上 $\alpha=5\%$的相应值时，称 x 与 y 没有明显的线性关系．在上例中 $r=0.9894$．因为 $n=10$，查表中 $\alpha=5\%$相应值为 0.632，$\alpha=1\%$相应值为 0.765，$r=0.9894>0.765$，故最大积雪深度与灌溉面积有十分显著的线性关系．显然当散点图出现明显的非线性趋势，或经检验，x 和 y 之间没有明显的线性关系时，再利用线性假设就不合理了．

§7.2　线性回归中基函数的选取

在数学分析中，常常将一个函数展开成级数，如泰勒级数、三角级数等，取级数的前 n 项，近似地表示这个函数．在回归分析中，也常常假定要估计的函数可以用某一类给定的函数展开成级数，取其前几项近似地表示成

$$y=c_1\varphi_1(x)+c_2\varphi_2(x)+\cdots+c_k\varphi_k(x), a\leqslant x\leqslant b$$

其中，$\varphi_j(k)$，$j=1,2,\cdots,k$ 是一组线性无关函数，它们一般是 x 的非线性函数，c_j，$j=1,2,\cdots,k$ 是回归系数．这个模型也称为线性回归模型，因为问题对回归系数而言是线性的．这个问题的更一般的提法是曲线拟合或函数逼近，当然逼近的方式并不局限于回归中的最小二乘法．其中 $\sum_{j=1}^{k} c_j\varphi_j(x)$ 称为 $y=f(x)$ 的线性逼近函数，c_j，$j=1,2,\cdots,K$ 称为线性逼近参数．

在曲线拟合问题中，常常采用半线性逼近函数类，即

$$y=F(\boldsymbol{c},\boldsymbol{d},x)=\varphi_0(\boldsymbol{d},x)+\sum_{j=1}^{n} c_j\varphi_j(\boldsymbol{d},x)$$

其中，$\boldsymbol{c}=(c,\cdots,c_n)$ 称为线性逼近参数，$\boldsymbol{d}=(d,\cdots,d_q)$ 称为非线性逼近参数．例如，当 $q=h$，$\varphi_0(\boldsymbol{d},x)\equiv 0$，$\varphi_j(\boldsymbol{d},x)=\exp(-d_j x)$，$j=1,2,\cdots,n$，则得到

$$y=\sum_{j=1}^{n} c_j e^{-d_j x}$$

这一函数类有重要应用．

下面是几个常用的半线性逼近函数类：

$y=a+be^{-cx}$，$y=ae^{-bx}$，

$y=a\cos(bx+c)$　或　$y=\alpha\cos rx+\beta\sin rx$

$y=a+bx+ce^{-dx}$，$y=(a+bx)e^{-cx}$．

这里 a、b、c、d 均为待定参数．

用 l_p 模来表示曲线拟合的误差，记

$$e(x)=f(x)-F(\boldsymbol{c},x)$$

$e(x)$ 的 l_p 模被定义为

$$\|e\|_p=\left\{\int_a^b \omega(x)|e(x)|^p dx\right\}^{\frac{1}{p}}$$

这里 $\omega(x)$ 称为权函数，$\omega(x)\equiv 0$ 且不恒为零，选取曲线模型的参数，使

$$\|e\|_p=\min$$

通常 p 取成 1,2 或 ∞. 如果 $p=2$,就是最小二乘意义逼近,即通常所说的回归,这在实际中应用最为广泛. 在这里研究 $p=2$ 的情况,但仍采用"拟合""逼近"的说法.

在曲线逼近中,常用的有插值法和最小二乘法,两者是不同的概念,但又有联系.

给定函数 $f(x)$ 的一组观测数据 (x_i,f_i), $i=1,2,\cdots,m$,选取 $\sum_{i=1}^{n}c_i\varphi_i(x)$ 为逼近函数,假定 m 维欧氏空间的向量组

$$\boldsymbol{x}_1=\begin{pmatrix}\varphi_1(x_1)\\ \varphi_1(x_2)\\ \cdots\\ \varphi_1(x_m)\end{pmatrix},\quad \boldsymbol{x}_2=\begin{pmatrix}\varphi_2(x_1)\\ \varphi_2(x_2)\\ \cdots\\ \varphi_2(x_m)\end{pmatrix},\quad \cdots,\quad \boldsymbol{x}_n=\begin{pmatrix}\varphi_n(x_1)\\ \varphi_n(x_2)\\ \cdots\\ \varphi_n(x_m)\end{pmatrix}$$

是线性无关的. 记

$$(f,g)=\sum_{i=1}^{m}w_if(x_i)g(x_i)$$

则所谓最小二乘逼近,就是选取 c_i,使

$$\sum_{k=1}^{m}w_k\left(f(x_k)-\sum_{i=1}^{n}c_i\varphi_i(x_k)\right)^2=\min$$

可证明问题的解是存在且唯一的.

现在,假定 $m=n$,选取 c_i 满足下列插值问题的解

$$\sum_{i=1}^{n}c_i\varphi_i(x_k)=f(x_k),\quad k=1,2,\cdots,n.$$

由于 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 是线性无关的,因而插值问题的解存在且唯一. 即能找到一组 c_i^*,使

$$\sum_{k=1}^{n}w_k\left(f(x_k)-\sum_{i=1}^{n}c_i^*\varphi_i(x_k)\right)^2=0$$

此即最小二乘形式,由最小二乘逼近解的存在唯一性,c_i^* 即为最小二乘的解. 这说明在 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 线性无关的假设下,最小二乘具有插值法的自适应性.

在曲线拟合中最重要的问题是选择基函数 $\{\varphi_i\}^n$,不同的问题

应有不同的选法，其重要依据是各学科的背景．随着电子计算机的发展，根据数据$(x_i,f_i)i=1,2,\cdots,m$的特征（散点图），利用专用数值软件和人工智能方法由计算机作出模型与方法的选择，是近年来国外较为重视的一个研究课题，属于建模支持系统的范围，这儿不作深入研究．

在曲线拟合中，代数多项式是一个基本函数类，取$\varphi_j(\boldsymbol{d}_1x)=x^j$，则

$$y=c_0+c_1x+c_2x^2+\cdots+c_nx^n$$

即在非线性拟合中采用的多项式逼近．这一基函数类虽然简单，但在很多场合并不恰当．这里介绍两类有用的基本函数类．

1. 正交多项式

在数学分析中，常常用正交函数系把一个函数展开成函数级数，即令回归模型为

$$f(x)=c_1h_1(x)+c_2h_2(x)+\cdots+c_kh_k(x)$$

选用这样的$h_i(x)$，使在观测点集合$\{x_t,y_t,t=1,2,\cdots,N\}$上满足

$$\begin{cases}\sum\limits_{t=1}^{N}h_i^2(x_t)>0, & i=1,2,\cdots,k\\ \sum\limits_{t=1}^{N}h_i(x_t)h_j(x_t)=0, & i\neq j\end{cases}$$

则有

$$\begin{cases}\overline{h}_i(x_t)=\dfrac{1}{N}\sum\limits_{t=1}^{N}h_i(x_t)=0\\ l_i=\sum\limits_{t=1}^{N}h_i^2(x_t)\\ l_{iy}=\sum\limits_{t=1}^{N}h_i(x_t)y_t\end{cases}$$

于是正规方程为

$$\begin{bmatrix} \sum_{i=1}^{N} h_1^2(x_i) & & & \\ & \sum_{i=1}^{N} h_2^2(x_i) & & \\ & & \ddots & \\ & & & \sum_{i=1}^{N} h_k^2(x_i) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \cdots \\ c_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} h_1(x_i) y_i \\ \sum_{i=1}^{N} h_2(x_i) y_i \\ \cdots \\ \sum_{i=1}^{N} h_k(x_i) y_i \end{bmatrix}$$

所以 $h_i(x)$ 的回归系数为

$$c_i = \sum_{i=1}^{N} h_i(x_i) y_i \Big/ \sum_{i=1}^{N} h_i^2(x_i) \qquad (i=1,2,\cdots,k).$$

常用的正交多项式有

2. 契比晓夫多项式

契比晓夫多项式被定义为

$$T_n(x) = \cos n\theta,\ 0 \leqslant \theta \leqslant \pi$$

$$\cos\theta = x,\ |x| \leqslant 1$$

选取契比晓夫多项式作基函数，不但具有最小二乘和插值法的自适应性，而且还有选取基函数个数的自适应性．利用契比晓夫多项式的正交性质，Shampine 于 1975 年指出了回归系数 c_k 的一个公式

$$c_k = \frac{(\varepsilon_{k-1}, T_k)}{\| T_k \|^2}$$

其中，$\varepsilon_{k-1}(x_l) = f(x_l) - \sum_{i=0}^{k-1} c_i T_i(x_l)$，表示拟合误差，由许瓦兹公式有

$$|c_k| \leqslant \sqrt{\frac{2\sum_{l=0}^{m} \varepsilon_{k-1}^2(x_l)}{m+1}}$$

因而，如果 $\sum_{i=0}^{k-1} c_i T_i(x)$ 的拟合效果好，c_k 的系数便趋于零，即由 $|c_k|$ 的大小可判别计算是否停止．

契比晓夫多项式的前 n 个为

$T_0=1,T_1=x,T_2=2x^2-1,T_3=4x^3-3x,\cdots,$

它具有递推关系：

$$\begin{cases}T_{n+1}(x)=2xT_n(x)-T_{n-1}(x) \quad n=1,2,\cdots \\ T_0(x)=1,T_1(x)=x\end{cases}$$

在利用契比晓夫多项式作拟合(或回归)时，须将观测数据中的 x_i，转换到$[-1,1]$区间上.

3. 勒让德多项式

$$P_0(x)=1,P_n(x)=\frac{1}{2^n n!}\frac{d^n}{dx^n}[(x^2-1)^n] \quad n=1,2,\cdots$$

它具有递推形式

$$\begin{cases}P_{k+1}(x)=\frac{2k+1}{k+1}xP_k(x)-\frac{k}{k+1}P_{k-1}(x) \quad k=1,2,\cdots \\ P_0(x)=1,P_1(x)=x.\end{cases}$$

其前 n 项为：$P_0(x)=1,P_1(x)=x,P_2(x)=\frac{1}{2}(3x^2-1),P_3(x)=\frac{1}{2}(5x^2-3x),\cdots.$

4. 埃尔米特多项式

$$H_n(x)=(-1)^n e^{x^2}\frac{d^n}{dx^n}(e^{-x^2})$$

它是在区间$(-\infty,\infty)$上带权 e^{-x^2}的 n 次正交多项式，具有递推公式

$$\begin{cases}H_{k+1}(x)=2xH_k(x)-2kH_{k-1}(x) \quad k=1,2,\cdots \\ H_0(x)=1,H_1(x)=2x\end{cases}$$

5. 拉盖尔多项式

$$L_n(x)=e^x\frac{d^n}{dx^n}(x^n e^{-x})$$

它是在$[0,\infty)$上带权 e^{-x}的 n 次正交多项式

具有递推公式

$$\begin{cases} L_{k+1}(x)=(1+2k-x)L_k(x)-k^2L_{k-1}(x) \quad k=1,2,\cdots \\ L_0(x)=1, L_1(x)=1-x \end{cases}$$

6. 指数函数

对于一类实验数据，它们之间呈现着指数并含有振荡的特征，在这种情况下，选取指数函数、三角函数、多项式为基函数是合适的．具体做法如下．

首先通过观测数据(k,f_k)，$k=0,1,\cdots,m$ 将首项系数为1的n次代数方程

$$p(\lambda)=\lambda^n+a_1\lambda^{n-1}+\cdots+a_{n-1}\lambda+a_n=0 \tag{1}$$

的系数a_i(均系实数)$i=1,2,\cdots,n$确定下来．这里$m\geqslant 2n-1$．类似于n阶常系数常微分方程的通解构造，称(1)为拟合曲线的特征方程，将通过根的不同情况作出基函数的不同选择．也即是说，将曲线拟合的特征方程(1)的根$\lambda_1,\lambda_2,\cdots,\lambda_n$的$x$次方程作为基函数，曲线模型选为

$$y=c_1\lambda_1^x+c_2\lambda_2^x+\cdots+c_n\lambda_n^x \tag{2}$$

下面针对不同特征根，类似于高阶常系数常微分方程通解的构造，将λ_i^x确定下来．

1° 特征根是单根情况

假定特征方程(1)有n个互异的根$\lambda_1,\lambda_2,\cdots,\lambda_n$．如果所有$\lambda_i>0$，那么，$\lambda_i^k=e^{x\ln\lambda_i}$；如果某一个$\lambda_i<0$，则

$$\lambda_i^x=(-1)^x|\lambda_i|^x=(\cos\pi x+i\sin\pi x)e^{x\ln|\lambda_i|}.$$

考虑到实数解，仅取实数部分，即取

$$\lambda_i^x=\cos\pi x e^{x|\ln\lambda_i|}$$

作为基函数；如果λ_i是复根，假定λ_{i+1}是它的共轭复根．

$$\lambda_i=\rho_i(\cos\theta_i+i\sin\theta_i)$$

类似于微分方程的讨论，将基函数取成λ_i^x的实数和复数部分，即

$$\lambda_i^x=e^{x\ln\rho_i}\cos\theta_i x, \lambda_{i+1}^x=e^{x\ln\rho_i}\sin\theta_i x$$

2° 方程(1)有k重根情况

假定$\lambda_i=\lambda_{i+1}=\cdots=\lambda_{i+k-1}\neq\lambda_{i+k}$，则将$\lambda_i^x,\cdots,\lambda_{i+k-1}^x$取成

$\lambda_i^x = \lambda_i^x \lambda_{i+1}^x = x\lambda_i^x, \cdots, \lambda_{i+k-1}^x = x(x-1)\cdots(x-k+2)\lambda_i^x$.

3° 方程(1)有零根

若 $\lambda_n = 0$，则将基函数减少一个，如果有 k 重零根，则将基函数减少 k 个．

现在，来确定特征方程(1)的系数．在(2)中，令 x 等于 $k+l$，这里 k、l 均为正整数．两端乘以 a_{n-k}，并对 k 从 0 到 n 求和，利用(1)，含 $y(k+l) = f_{k+l}, a_0 = 1$，便有

$$\sum_{k=0}^{k} f_{k+l} a_{n-k} = 0, l = 0, 1, \cdots, n$$

于是 $a_1, a_2, \cdots, a_n$ 由下列线性代数方程组决定：

$$\begin{cases} f_{n-1}a_1 + f_{n-2}a_2 + \cdots + f_0 a_n = -f_n \\ f_n a_1 + f_{n+1} a_2 + \cdots + f_1 a_n = -f_{n+1} \\ \cdots \\ f_{m-1}a_1 + f_{m-2}a_2 + \cdots + f_{m-n} a_n = -f_m \end{cases}$$

写成矩阵形式有

$$A\boldsymbol{x} = \boldsymbol{b}$$

这里

$$A = \begin{bmatrix} f_0 & f_1 & \cdots & f_{n-1} \\ f_1 & f_2 & \cdots & f_n \\ \cdots & \cdots & \cdots & \cdots \\ f_{m-n} & f_{m-n+1} & \cdots & f_{m-1} \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} a_n \\ a_{n-1} \\ \vdots \\ a_1 \end{bmatrix}, \boldsymbol{b} = -\begin{bmatrix} f_n \\ f_{n+1} \\ \vdots \\ f_m \end{bmatrix}$$

若 $m = 2n-1$，且 $\det A \neq 0$，则可用 Gauss 消去法求得 $\boldsymbol{x}$ 的解；若 $m > 2n-1$，则可用最小二乘法求解．

$\lambda_1, \lambda_2, \cdots, \lambda_n$ 确定后，通过(2)将拟合曲线的参数 $c_1, c_2, \cdots, c_n$ 确定下来：

$$\begin{cases} c_1 + c_2 + \cdots + c_n = f_0 \\ c_1\lambda_1 + c_2\lambda_2 + \cdots + c_n\lambda_n = f_1 \\ \cdots \\ c_1\lambda_1^m + c_2\lambda_2^m + \cdots + c_n\lambda_n^m = f_m \end{cases}$$

由于 $m \geqslant 2n-1$，故宜采用最小二乘法求解．

上述处理方法称为 Prony 技巧，它将半线性逼近问题转化为两个线性逼近问题．第一步，利用已知数据选择好基函数，这一步包括解一个线性方程组和一个高次方程．第二步，再次利用观测数据将线性逼近参数 $c_1, c_2, \cdots, c_n$ 确定下来．

下面，利用 Prony 技巧研究两种特殊情况．

第 1 种情况，当 $x \to \infty$ 时，$f(x)$ 趋向定值．这要求拟合曲线也具有渐近性．这时，曲线拟合模型取成

$$Y = c_0 + c_1 \lambda_1^x + \cdots + c_n \lambda_n^x \tag{3}$$

记 $\Delta Y = y(x+1) - y(x)$，则可将上述模型化成

$$\Delta Y = c_1' \lambda_1^x + c_2' \lambda_2^x + \cdots + c_n' \lambda_n^x \tag{4}$$

在 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 确定后，由(1)式确定 $c_1, c_2, \cdots, c_n$.

第 2 种情况，如函数 $f(x)$ 由不同的周期 $\frac{2\pi}{\omega_1}, \cdots, \frac{2\pi}{\omega_p}$ 迭加而成，则将拟合曲线取成

$$y = A_1 \cos\omega_1 x + B_1 \sin\omega_1 x + \cdots + A_l \cos\omega_l x + B_l \sin\omega_l x.$$

应用 Prony 技巧，$\lambda = e^{i\omega_j}$，在(1)式中令 $n = 2l$，注意到 $\frac{1}{\lambda}$ 也是根，因而 $a_{2l-1} = a_1, \cdots, a_{l+1} = a_{l-1}, a_{2l} = 1$，故

$$e^{i2l\omega_j} + a_1 e^{i(2l-1)\omega_j}$$

$$+ \cdots + a_{l-1} e^{i(l-1)\omega_j} + a_l e^{il\omega_j} + a_{l-1} e^{i(l-1)\omega_j} + \cdots + a_1 e^{i\omega_j} + 1 = 0$$

合并同类项整理得

$$2\cos l\omega_j + 2a_1 \cos(l-1)\omega_j + \cdots + 2a_{l-1} \cos\omega_j + a_l = 0$$

这时(1)可写成契比晓夫多项式形式

$$T_l(x) + a_1 T_{l-1}(x) + \cdots + a_{l-1} T_1(x) + \frac{1}{2} a_l = 0$$

其中，$x = \cos\omega$.

而方程组变成

$$
\begin{cases}
(f_1+f_{2l-1})a_1+(f_2+f_{2l-2})a_2+\cdots+(f_{l-1}+f_{l+1})a_{l-1} \\
+f_la_l=-(f_0+f_{2l}) \\
(f_2+f_{2l})a_1+(f_3+f_{2l-1})a_2-\cdots+(f_l+f_{l+2})a_{l-1} \\
+f_{l+1}a_l=-(f_1+f_{2l-1}) \\
\qquad\cdots \\
(f_{m-2l+1}+f_{m-1})a_1+(f_{m-2l+2}+f_{m-2})a_2+\cdots+(f_{m-l-1} \\
+f_{m-l+1})a_{l-1}+f_{m-l}a_l=-(f_{m-2l}+f_m).
\end{cases}
$$

这里，$m\geqslant 3l+1$. 进一步，利用通常的方法将线性逼近参数 A_i, B_i 确定下来.

例 7-2 观测数据表 7-2 所示，试研究其规律

表 7-2

x	0	1	2	3	4
$f(x)$	2.4400	2.0851	2.1958	2.2692	2.3006

在运算中保留小数点后两位数字，求得特征方程

$$183\lambda^2-91\lambda+8=0$$

其根为 $\lambda_1\approx 0.383, \lambda_2\approx 0.114$，从而拟合模型为

$$
\begin{aligned}
y=F(c,x)&=c_0+c_1(0.383)^x+c_2(0.114)^x \\
&=c_0+c_1e^{-0.96x}+c_2e^{-2.18x}
\end{aligned}
$$

再按通常线性逼近的方法，将参数 c_0, c_1, c_2 确定下来. 实际上观测数据取自函数

$$f(x)=2.32-1.08e^{-x}+1.20e^{-2x}$$

这说明拟合模型的基函数有近似再生性.

例 7-3 某镇的用水机构需估计公众用水速度（单位：加仑/小时）和每日总用水量的数据. 许多地方没有测量流入或流出市政水箱流量的设备而只能测量水箱中的水位（误差不超过 0.5%），当水箱水位低于最低水位 L 时，水泵抽水灌入水箱直到水箱水位达到最高水位 H 为止，但是无法测量水泵的流量，因此，

在水泵开动时无法立即将水箱中的水位和用水量联系起来．这种情形一天发生一次或两次，每次约为两小时．

估计所有时刻，包括水泵抽水期间流出水箱的流量 $f(t)$，并估计一天总用水量．附表给出某一天某小镇的真实数据．

表中以秒为单位给出距开始测量的时间，水位单位是 1%公尺．水箱为高 40 公尺、直径 57 公尺的正圆柱，通常水泵当水位落到 27 公尺以下开始抽水而当水位回升至 35.5 公尺时，水泵停止工作．

表 7-3　水箱水位观测数据

时间(秒)	水位(0.01 公尺)	时间(秒)	水位(0.01 公尺)
0	3175	46636	3350
3316	3110	49953	3260
6635	3054	55936	3167
10619	2994	57254	3087
13937	2947	60574	3012
17921	2892	64554	2927
21240	2850	68535	2842
25223	2795	71854	2767
28543	2752	75021	2697
32284	2697	79254	水泵开动
35932	水泵开动	82649	水泵开动
39332	水泵开动	85968	3475
39435	3550	89953	3397
43318	3445	93270	3340

在建模过程中假定：

1° 各个测量值没有系统误差，而只有偶然误差即由于观测等原因带来的误差．

2° 水流速率是连续变化的，且不出现突起突落，因为对于许多用户来说用水情况表现出一定的规律性，某个随机因素对总个用水的情况不会影响很大．

3° 不同两天的相同时刻用水情况相近，即水流速率表现周期性．

记 F 为水箱流量，H 为水箱水位，则在 $\Delta t=(t+\Delta t)-t$(秒)内，水位变化值 ΔH(0.1 公尺)和水流量变化值 ΔF 之间满足

$$\Delta F=\pi*(\frac{D}{2})^2\times\Delta H\times\frac{(12\times2.539998)^3}{3.7853}$$
$$=190.89\Delta H$$

两边同除以 Δt，令 $\Delta t\to0$，得

$$f(t)=190.89h(t)$$

这表明水流速度和水箱水位变化率为线性关系．下面着重分析水位变化率．

设 t_i 时刻的水位高度为 H_i，t_{i-1}时刻的高度为 H_{i+1}，则在 $t_i\sim t_{i-1}$的平均变化率为 $h_i=\frac{H_{i+1}-H_i}{t_{i+1}-t_i}$，这样 $h(t)$可近似表示为阶梯函数．取各小时间段中点，利用最小二乘进行曲线拟合，求取 $h(t)$.

用线段连接各时间段 $h(t)$的中点，知 $h(t)$为多峰值，若采用一般多项式回归，阶数较高，并且回归系数间存在相关性．考虑选取基函数的方法，例如取正交多项式．

设 $h(t)=d_0h_0(x)+d_1h_1(x)+\cdots+d_ih_i(x)+\cdots+d_kh_k(x)$

问题转为求正交多项式．但这要求待拟合的各点等间距，采用线性插值补齐等间距点．对－12 小时到＋12 小时的数据进行拟合，得

$$h(t)=71.225-4.516t-0.680t^2+0.0504t^3+0.0126t^4$$
$$-0.000132t^5-4.521\times10^{-5}t^6,$$

$$-12\leqslant t\leqslant 12（小时）$$

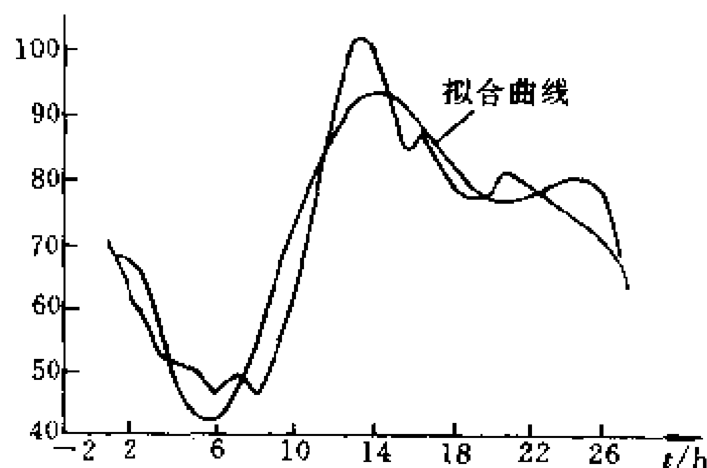
这时残差平方和为 170.227（若用 7 阶多项式，残差平方总和为 634.58），在 -12 小时和 $+12$ 小时时刻水位变化率分别为 99.76 和 99.64，基本上保持了连续．

故水流速度函数为

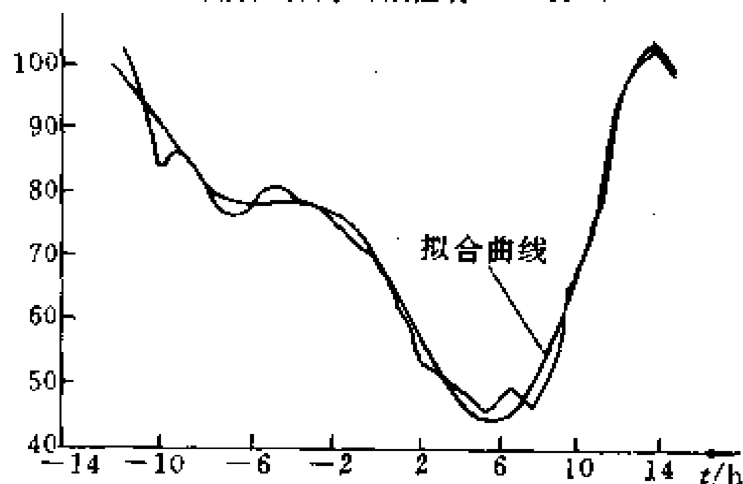
$$f(t)=190.89\times(71.225-4.516t-0.680t^2+0.0504t^3 +0.0126t^4-0.000132t^5-4.521\times10^{-5}t^6)$$

一天总水位变化为

$$H=\int_{-12}^{12}h(t)\mathrm{d}t=1721.43$$



（a）直接拟合的函数图像（0～25 h）



（b）处理后的拟合图像（−12～12h）

图 7-2

一天中总用水量为

$$T=190.89\times H=3.286\times 10^5$$

对水位的测量值和预测值的绝对误差作统计检验，服从 $N(0.112,2.94^2)$，误差值和测量值相关系数为 0.0943，相对误差绝对值的平均值为 3.92%，说明用正交多项式拟合的性能良好．

§7.3 曲线拟合和常微分方程反问题

曲线拟合的另一方法是将基函数看成是某一个待定参数的常微分方程的基解组，那么，问题便成为求解常微分方程的反问题．一般讲来，物理、力学所描述的规律由微分方程所表示，通过观测数据将微分方程的参数确定下来，进而将基解组作为曲线拟合的基函数．最后，运用线性逼近方法将拟合曲线确定下来．这个方法的一个重要优点是有一定的自适应性，从数学上讲，它是解决半线性逼近的重要方法．

我们从一个实际问题分析起，在实验数据拟合中，常常选择半线性逼近模型，例如对具有指数特征的数据，选取

$$y=a\mathrm{e}^{kx}+b\mathrm{e}^{-kx},$$

对于具有周期特征的数据，选取

$$y=a\cos kx+b\sin kx,$$

上述模型的半线性逼近参数 k 通常采用非线性最小二乘法确定下来．但运算量很大．现在，将 e^{kx}，e^{-kx} 以及 $\cos kx$，$\sin kx$ 理解为二阶常系数微分方程的基解组，即

$$y''+\alpha y=0,$$

其中 α 是待定常数，它由观测数据 $(x_i,f_i)i=0,1,\cdots,m+1$ 所确定．这里 $y=f(x)$ 是 $y''+\alpha y=0$ 的解，记 $y_i=f(x_i)=f_i$，x_i 是等距分布的，步长为 h，即 $x_i=a+ih$．也就是说，求解下列常微分方程的反问题：

$$\begin{cases}y''-\alpha y=0, & a\leqslant x\leqslant b\\ y(x_i)=f(x_i), & i=0,1,\cdots,m+1\end{cases}$$

将微分方程离散化后，α 便由下列超定方程组所确定

$$\alpha h^2 y_i = -\Delta^2 y_{i-1}, i=1,2,\cdots,m$$

这里 $\Delta^2 y_{i-1} = y_{i+1} - 2y_i + y_{i-1}$，用最小二乘法求解，即要求 α，使

$$\min I(\alpha) = \sum_{i=1}^{m} (\alpha h^2 y_i + \Delta^2 y_{i-1})^2$$

令 $I'(\alpha)=0$，求得

$$\alpha = -\sum_{i=1}^{m} y_i \Delta^2 y_{i-1} \Big/ h^2 \sum_{i=1}^{m} y_i^2$$

显然，当 $y=f(x)$是线性函数时，有 $\Delta^2 y_{i-1} \equiv 0$ 故 $\alpha=0$ 从而微分方程变成 $y''=0$，即 $y=ax+b$，即反解问题对线性函数具有再生性．如果

$$y=f(x)=a\cos kx + b\sin kx$$

那么，由于 α 的分母恒正，而当 h 充分小时，

$$-\sum_{i=1}^{m} y_i \Delta^2 y_{i-1}/h \approx -\int_a^b y y'' \mathrm{d}x = k^2 \int_a^b y^2 \mathrm{d}x > 0$$

故 $\alpha>0$，从而 $y''+\alpha y=0$ 的特征根是复共轭的，或者说反解问题具有近似再生性质，而当曲线

$$y=f(x)=a\mathrm{e}^{kx}+b\mathrm{e}^{-kx}$$

时，由于

$$-\sum_{i=1}^{m} y_i \Delta^2 y_{i-1}/h = -\int_a^b y y'' \mathrm{d}x = -k^2 \int_a^b y^2 \mathrm{d}x$$

故 $\alpha<0$，从而反解问题对指数曲线也具有近似再生性质．实际上，当 h 充分小时

$$\begin{aligned}\alpha &= -\sum_{i=1}^{m} y_i \Delta^2 y_{i-1}/h^2 \sum_{i=1}^{m} y_i^2 \approx -\int_a^b y y'' \mathrm{d}x \Big/ \int_a^b y^2 \mathrm{d}x \\ &\approx \pm k^2.\end{aligned}$$

式中，若 $f(x)=a\mathrm{e}^{kx}+b\mathrm{e}^{-kx}$，则取负号，若 $f(x)=a\cos kx+b\sin kx$ 则取正号．注意到求积公式的准确度，容易看到 α 逼近 $\pm k^2$ 的误差为 $o(\lambda)$. 这一结果可利用 Peano 核定理得到一般性证明．

例 7-4 给定函数表 7-4 试找出拟合曲线的基函数．

表 7-4 函数数值表

x	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
$f(x)$	0.0000	0.10017	0.20134	0.30452	0.41075	0.52110	0.63665	0.75858

为确定反解问题的系数 α，作下列二次差分表 7-5.

表 7-5 二次差分表

x	y_i	y_i^2	Δy_{i-1}	$\Delta^2 y_{i-1}$	$y_i\Delta^2 y_{i-1}$
0	0.0000				
0.1	0.10017	0.010034	0.10017	0.00100	0.010017×10^{-2}
0.2	0.20134	0.405377	0.10117	0.00201	0.0404693×10^{-2}
0.3	0.30452	0.0927324	0.10318	0.00305	0.0928786×10^{-2}
0.4	0.41075	0.1687155	0.10623	0.00412	0.169229×10^{-2}
0.5	0.52110	0.2715452	0.11035	0.00520	0.270972×10^{-2}
0.6	0.63665	0.4053232	0.11555	0.00638	0.4061827×10^{-2}
0.7	0.75858		0.12193		
$\sum$	0.988888				0.9897486×10^{-2}

这里 $h=0.1$，所以有

$$\alpha=-\sum_{i=1}^{6}y_i\Delta^2 y_{i-1}/h^2\sum_{i=1}^{6}y_i^2=-\frac{0.9897486}{0.988888}=-1.0008702$$

于是可取 $e^{\sqrt{|\alpha|}x}$，$e^{-\sqrt{|\alpha|}x}$为拟合曲线的基，而 $\sqrt{|\alpha|}=1.0004$，即拟合模型为

$$y=c_1e^{1.0004x}+c_2e^{-1.0004x}.$$

实际上，函数表是由双曲正弦函数 $y=\frac{1}{2}(e^x+e^{-x})$产生的．可见，利用反问题求得的基函数确有近似再生的特征．

利用反问题求曲线拟合的基函数，其最大优点是具有近似再

生性．其局限在于要求 x_i 为等距分布，即 $x_i=a+ih$，h 为步长，这一点需要在试验设计时考虑到．

§7.4 多元回归与曲面拟合

设对 y 及 $x_1,x_2,\cdots,x_N$ 作 M 次观测后，得到一组观测值

$$(y_i,x_{i1},x_{i2},\cdots,x_{iN}),\quad i=1,2,\cdots,M$$

希望能从中找到 y 与诸 x_i 的关系式，即

$$y=f(x_1,x_2,\cdots,x_N).$$

如果已找到适当的曲线模型

$$y=F(\boldsymbol{c},\boldsymbol{x}),$$

则问题转化为利用观测数据对模型参数 $\boldsymbol{c}=(c_1,c_2,\cdots,c_m)$进行估计，要求选取的 $\boldsymbol{c}$ 使

$$\min\sum_{i=1}^{M}[y_i-F(\boldsymbol{c},\boldsymbol{x}_i)]^2.$$

对这一非线性规划问题，可采用逐次线性化和变尺度等方法求解.

特别地，当选取的曲线模型为线性函数时，即

$$y=\beta_0+\beta_1x_1+\cdots+\beta_Nx_N$$

若 b_j 分别为 $\beta_j(j=0,1,\cdots,N)$的估计值，用最小二乘法求 b_j，即

$$\min Q=\sum_{i=1}^{M}(y_i-b_0-b_1x_{i1}-\cdots-b_Nx_{iN})^2$$

由$\dfrac{\partial Q}{\partial b_j}=0(j=0,1,\cdots,N)$，得

$$\begin{cases}s_{11}b_1+\cdots+s_{N1}b_N=s_{y1}\\ s_{1N}b_1+\cdots+s_{NN}b_N=s_{yN}\end{cases}\tag{1}$$

其中

$$s_{ij}=\sum_{k=1}^{M}(x_{ik}-\bar{x}_i)(x_{jk}-\bar{x}_j)\quad(i,j=1,2,\cdots,N)$$

$$s_{yj}=\sum_{k=1}^{M}(y_k-\bar{y})(x_{jk}-\bar{x}_j)\quad(j=1,2,\cdots,N)$$

$$\bar{x}_i=\frac{1}{M}\sum_{k=1}^{M}x_{ik},\quad \bar{y}=\frac{1}{M}\sum_{k=1}^{M}y_k$$

(1)式被称为正规方程组，由此解得 $b_j, j=1,2,\cdots,N$.

对多元问题，模型函数的选取具有决定性意义．对二元函数，在选取基函数时，有几条原则可供参考．

1. 数据分布

对分布在矩形网格点上的数据，应用乘积型方法去逼近．若数据分布是散乱的，常用的方法有样条函数的最小二乘法，Shepard方法和二步逼近法等（参见文献[24]）.

2. 数据的准确度

在大量实际问题中，实测数据仅有三至四位准确，一般来讲，在这种情况下，局部逼近优于全局逼近．所谓局部逼近即"分片法"，也就是将大范围的拟合问题分割为许多互不相交的小片上的拟合，方法的优点是简单，容易保持几何特性，缺点是在连接处甚至可能不连续．

3. 全局逼近、局部逼近和二步法逼近

全局逼近和局部逼近是分别在大范围和小范围内选择基函数．而在二次法逼近中，首先将非规则数据规则化，即将矩形的格子点的函数值补齐，然后利用乘积型方法将曲面构造出来，二次逼近法常用于非规则数据．

4. 几种常用基函数

基函数的选取是一个极重要而又不容易解决的问题，通常考虑的有：

乘积型基函数 $\{\varphi_i(x)\psi_j(y)\}$．例如可选取 $\{1,x,y,xy,x^2,y^2\}$．这就是通常采用的二元二次函数

$$z=a_0+a_1x+a_2y+a_3xy+a_4x^2+a_5y^2$$

作为模型函数．乘积型基函数的另外选取方法为在两个方向上采用 B 样条函数 $\{B_{i,k}(x)B_{j,k}(y)\}$，即在 x,y 方向上都采用局部逼近；或者在 y 方向上选 $1,y,y^2$ 为基（全局二次函数逼近）x 方向上

用 B 样条为基(局部样条逼近),即

$$z=F(x,y)=\sum_{m=0}^{2}\sum_{l=1}^{n}c_{lm}B_{l,4}(x)y^{m}$$

当考虑散乱数据逼近时,距离是一个重要因素,也就是说远离(x_i,y_i)的那些点的函数值对 $f(x_i,y_i)$的影响是微小的,因此基函数中常含有

$$\frac{1}{x_i(x,y)}=\frac{1}{\sqrt{(x-x_i)^2+(y-y_i)^2}}.$$

的因子,例如

$$z=F(x,y)=\sum_{i=1}^{N}a_ir_i^2(x,y)\log r_i(x,y)+b_1x+b_2y+b_3$$

Duchon. J 称之为箔板样条．或

$$z=F(x,y)=\sum_{i=1}^{N}a_ir_i^3(x,y)+b_1x+b_2y+b_3$$

文献中称之为准三次样条．

Lorente 于 1966 年指出一个结果,对构造基函数和选取基函数的个数具有指导意义:

如果 $f(x,y)$在$[0,1]\times[0,1]$上连续,那么存在着两组连续函数 $\varphi_i(x)$和 $\psi_i(x)$,$i=1,2,\cdots,5$,使 $f(x,y)$可表成

$$f(x,y)=\sum_{i=1}^{5}\theta_i(\varphi_i(x)+\psi_i(y))$$

这里 $\theta_i(\cdot)$是依赖于 $f(x,y)$的单变量连续函数．

例 7-5 在土豆生长期间,施用不同量的氮(N)和钾(K)肥量,产量结果如表 7-6 所示. 磷肥均固定在 $p=200$kg/ha 水平上.

一般可假定产量 η 是施肥量 N、K 的二次函数(这点从对 N 和 K 单因素的散点图上可以看到). 不直接将 η 表成 N,K 的函数,而是作中心化处理,引进新的变量

$$u_1=\frac{N-260}{60},\quad u_2=\frac{K-370}{100}$$

将 η 表为 u_1,u_2 的二次多项式．可以直接证明

表 7-6 施肥量与产量观测数据

序号	N(kg/ha)	K(kg/ha)	产量 y(T/ha)
1	200	270	48
2	200	370	37
3	200	470	49
4	260	270	44
5	260	370	42
6	260	470	52
7	320	270	49
8	320	370	43
9	320	470	59

$$\varphi_1(u)=1,\varphi_2(u)=u_1,\varphi_3(u)=u_1^2-\frac{2}{3},\varphi_4(u)=u_2,$$

$$\varphi_5(u)=u_2-\frac{2}{3},\varphi_6(u)=u_1u_2$$

在试验的 η 点上正交．可将二次模型表为二次正交多项式 $\eta(u)=\beta_1+\beta_2u_1+\beta_3(u_1^2-\frac{2}{3})+\beta_4u_2+\beta_5(u_2^2-\frac{2}{3})+\beta_6u_1u_2$，利用正交性，可得回归系数

$$\hat{\beta}_1=47,\hat{\beta}_2=2.8,\hat{\beta}_3=\frac{3}{2},\hat{\beta}_4=3.2,\hat{\beta}_5=9.5,\hat{\beta}_6=2.25.$$

在本例模型选择的讨论中，剔除和引进变量的问题，参见文献[2][18].

例 7-6 枪炮射表问题[27].

枪炮射表是一张以数值形式给出的二元函数表，自变量为水平距离 d 及垂直距离 H，函数为高角 α. 它指出，如欲击中水平距离为 d，垂直高度为 H 的目标，为了克服重力的影响，枪炮管的仰角应比目标的高低角 $\varepsilon=\sin^{-1}\frac{H}{\sqrt{d^2+H^2}}$高出多少，如图 7-3 所示.

由于空气阻力与重力的交互作用，在客观上函数 $\alpha(d,H)$不

能由理论力学算出，它的形式异常复杂．如果将表函数 $\alpha(d,H)$ 由实验数据整理出来，数据量很大，不能存入计算机．因此希望能找到它的一个近似的解析表达式，当然要求有较高的精度．

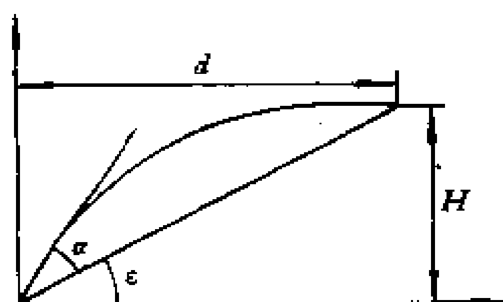


图 7-3　高角 α 与高低角 ϵ

显然，对于每个具体型号的枪炮而言，函数 $\alpha(d,H)$ 都是充分光滑的．因此，根据函数逼近理论中的 Weierstrass 定理，可取函数级数

$$\sum_{i,j=0}^{\infty} c_{ij} d^i H^j \tag{2}$$

中的有限项作为函数的近似解表达式，其中系数按最小二乘法，由表函数 $\alpha(d,H)$ 的数据得出．然而在实际计算发现，表函数 $\alpha(d,H)$ 的形式十分复杂，而且在 d 和 H 的数据跨度很大的情况下，舍入误差的存在使得考虑的问题不能实现．一个仅具有理论上的完备性的基函数系，在逼近中需要很多项才能达到必要的精度，而项数增多，则阶数必高，使数据的跨度变得越来越大，于是舍入误差的积累变得越来越严重，以致在远远还未达到必要的逼近精度以前，无论再怎样添加(1)式中的函数项，逼近的精度已不能再有所提高了．

为了改变这种状况，有如下两条出路：1° 分片逼近（这样做肯定可以提高逼近的精度，但是所得到的表达式在使用时计算过程变得复杂，同时需要事先存入计算机中的数据量也相应增加）；2° 分析表函数 $\alpha(d,H)$ 的性质，寻求较为贴切的逼近基函数系以提高逼近的效率．这种方案一般比较困难．寻求这种“贴切”的基函数需要有很好的针对性和相当的技巧．然而有时也还是有若干方法能有助于实现这个方案，这种方案一旦实现，效果必然会是相当理想的．其中一种方法就是分析同类数学模型中普遍存在的光

滑性、对称性、周期性、相似性、守恒性等，然后利用这些分析出来的性质，对基函数系加以自然的选择．

分析一张具体的表函数 $\alpha(d,H)$，其中 $d,H>0$，很难得出什么有用的性质．为了便于分析，不局限于直接考虑这个数值表函数，而先考虑这个数值表函数所要近似表现的那个客观存在的在力学上具有精确意义的函数 $\bar{\alpha}(d,H)$．为进一步分析 $\bar{\alpha}(d,H)$的性质，想象将 d,H 开拓成可取正值亦可取负值．这种延拓，从实战角度看，当然毫无意义，因不能旋转炮口，使其从瞄准前方向仰过头去瞄准后方，更不能使炮口朝着地下．但是，在抛开具体的技术困难后，从纯粹的数学和力学的角度考虑，这种延拓具有确切含义．如图 7-4 所示．

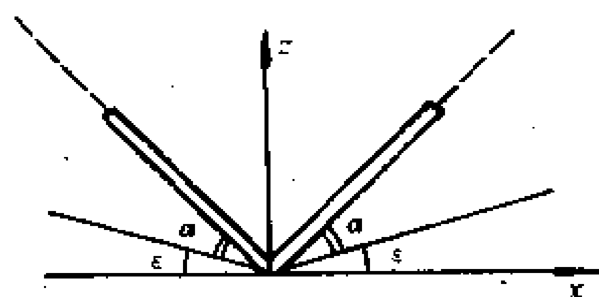


图 7-4　表函数 $\alpha(d,H)$的延拓

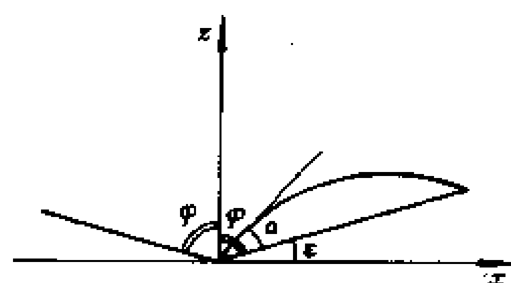


图 7-5　φ 角的图示

将延拓后的 d 及 H 分别记为 x 和 z，则原来的表函数 $\bar{\alpha}(d,H)$就是延拓后的表函数 $\bar{\alpha}(x,z)$在第一象限的一张数值表．

对于函数 $\bar{\alpha}(x,z)$，容易分析出它的周期性与对称性，同时它们仍然保持充分的光滑性．利用这些性质，即可选送一列较为"贴切"的逼近基函数．

为便于分析，在(x,z)平面上引进变量 D、φ（见图 7-5），其中 $D=\sqrt{x^2+z^2}$，即目标至炮位的距离，φ 为从 z 轴方向右旋至目标矢径的角度．在此参数系下，$\bar{\alpha}$ 为欲击中目标，炮管须自目标矢径方向左旋的角度．

在这样的坐标系与参量的选取之下，可写出高度函数 $\bar{\alpha}$ 的反

对称性质：

$$\bar{\alpha}(D,\varphi)=-\bar{\alpha}(D,-\varphi) \tag{3}$$

这种反对称性质的根源在于重力场是均匀的且其方向沿着 x 轴的相反方向．

还可以写出高度角函数 $\bar{\alpha}$ 的周期性

$$\bar{\alpha}(D,\varphi)=\alpha(D,\varphi+2\pi) \tag{4}$$

此周期性的根源纯粹基于参数系统的合适选取．

由(2)(3)式，再利用 Fourier 理论，可得展式

$$\bar{\alpha}(D,\varphi)=\sum_{n=1}^{\infty}a_n(D)\cdot\sin n\varphi \tag{5}$$

其中 $a_n(D)$表示 D的一元函数．

引进高低角 $\varepsilon=\varphi+\frac{\pi}{2}$，其几何意义为自 x 轴正向左旋目标矢径的角度，认为 $\bar{\alpha}$ 是 D、ε 的函数，则由(4)式得

$$\bar{\alpha}(D,\varepsilon)=\sum_{i=1}^{\infty}[b_{2i-1}(D)\cos(2i-1)\varepsilon+b_{2i}(D)\sin 2i\varepsilon] \tag{6}$$

其中 $b_{2i-1}(D)$，$b_{2i}(D)$表示 D 的一元函数．

注意到(5)式中的 $\cos(2i-1)\varepsilon$ 与 $\sin 2i\varepsilon$ 都可表示为如下形式：

$$\cos\varepsilon\cdot[c_0+\sum_{j=1}^{n}c_j\sin^j\varepsilon]$$

其中，n 为正整数，c_0，c_j 为实的绝对常数．

再利用 Weierstrass 逼近定理，将(5)式中的 $b_{2i-1}(D)$与$b_{2i}(D)$写成 D 的级数形式，于是得出

$$\bar{\alpha}(D,\varepsilon)=\cos\varepsilon\cdot\sum_{i,j=0}^{\infty}c_{ij}D^i\sin^j\varepsilon \tag{7}$$

其中，c_{ij}为实常数．于是逼近射表的基函数列可选为

$$\cos\varepsilon\cdot D^i\cdot\sin^j\varepsilon,\ i,j=0,1,2,\cdots \tag{8}$$

将上述讨论局限在第一象限，由定义

$$D=\sqrt{d^2+H^2},\quad \cos\varepsilon=\frac{d}{\sqrt{d^2+H^2}},\quad \sin\varepsilon=\frac{H}{\sqrt{d^2+H^2}}$$

于是所选基函数列为

$$dH^j/(d^2+H^2)^{\frac{1}{2}(j+1-i)}, i,j=0,1,2 \tag{9}$$

实际计算表明,这是一列效率相当高的基函数,它比(1)式中的二元多项式函数列要好得多,在所得的整体逼近的解析表达式中,函数项成几倍地减少,精度更是几十倍地提高.

对前苏联生产的卡斯-19式100mm口径高射炮射表作试验,基函数(8)式中的如下8项(如表7-7)所得结果中,函数 α 逼近误差的均方根为0.225. 其中 α 的单位为密位(1密位$=\frac{2\pi}{6000}$弧度),射表中 α 的最小刻度为1密位.

表7-7 基函数中的8项

i	1	2	3	4	1	2	1	2
j	0	0	0	0	1	1	2	2

§7.5 非线性回归

在很多实际问题中,当从散点图或从机理分析判断两变量之间的关系不是线性关系时,应考虑采用曲线拟合. 由于非线性描述的内容极为丰富,因此采用什么函数(或称模型)来作为拟合曲线,是非线性回归中的关键问题.

1. 内线性模型

有些模型,例如

$$y=\beta_0+\beta_1 e^x, y=\beta_0+\beta_1 \ln x, y=\beta_0+\beta_1 x^2$$

等,对自变量 x 都不是线性的,但 y 对参数 β_0 和 β_1 而言是线性的. 这时可以用适当的代换化为线性模型.

另外有些模型,例如

$$y=\frac{1}{\beta_0+\beta_1 e^x}, y=\beta_0 e^{\beta_1 x}$$

$$y=\beta_0 x^{\beta_1}, y=(\beta_0+\beta_1 x)^{-1}$$

等，虽然 y 对 x，对参数 β_0,β_1 都不是线性的，但也可以找到适当的变换，化为线性模型．

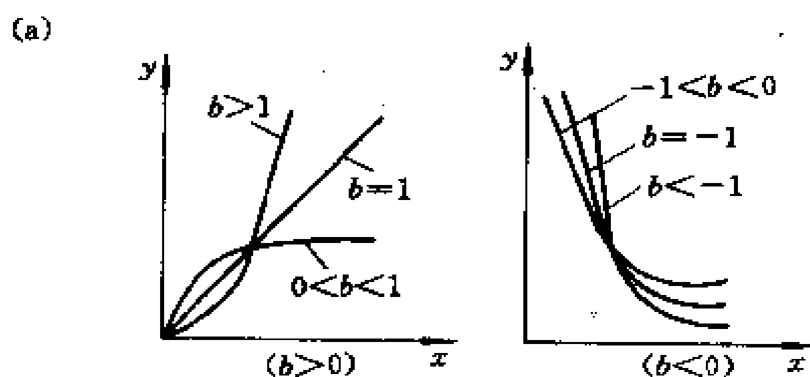
以上两类模型，统称为内线性模型，或称为可化为线性模型的回归问题．对这种问题的处理，一般可先化为线性模型，然后再用最小二乘法求出参数的估计值，最后再经过适当的变换，得到所求的回归曲线．这样经过变换再求回归曲线的方法，当然对估计参数的性质会产生影响，比如，不再具有无偏性等等．

六种常用的内线性模型及其变换列表 7-8.

表 7-8

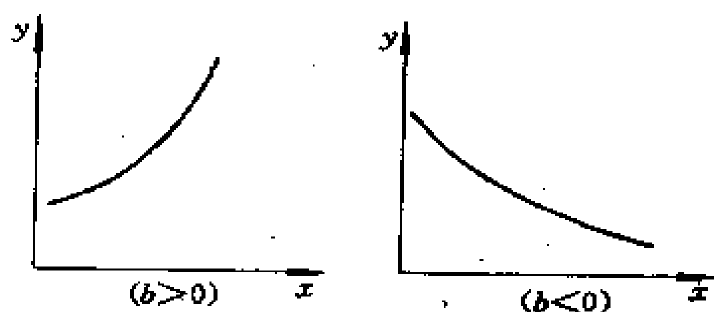
曲线	变换	变换后的线性式
幂函数 $y=ax^b$	$y'=\ln y, x'=\ln x$	$y'=\ln a+bx'$
指数函数 $y=ae^{bx}$	$y'=\ln y$	$y'=\ln a+bx'$
双曲函数 $y=\frac{x}{ax+b}$	$y'=\frac{1}{y}, x'=\frac{1}{x}$	$y'=a+bx'$
对数函数 $y=a+b\ln x$	$x'=\ln x$	$y'=a+bx'$
指数函数 $y=ae^{\frac{b}{x}}$	$y'=\ln y, x'=\frac{1}{x}$	$y'=\ln a+bx'$
S 型曲线 $y=\frac{1}{a+be^{-x}}$	$y'=\frac{1}{y}, x'=e^{-x}$	$y'=a+bx'$

这些非线性曲线的图形如图 7-6 所示．



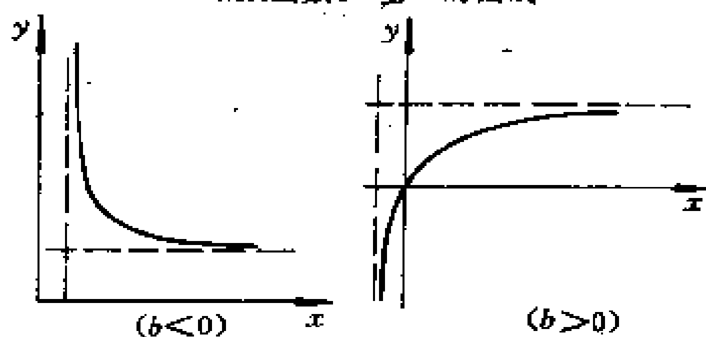
幂函数 $y=dx^b$ 的曲线

(b)



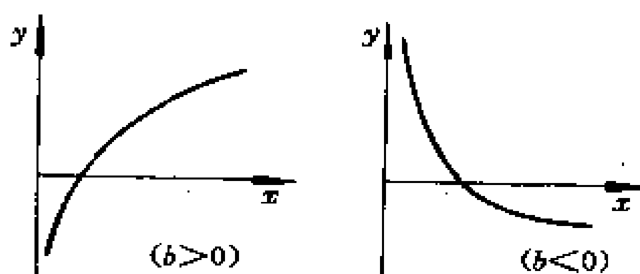
指数函数 $y=ae^{bx}$的曲线

(c)

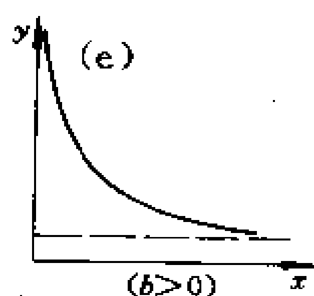


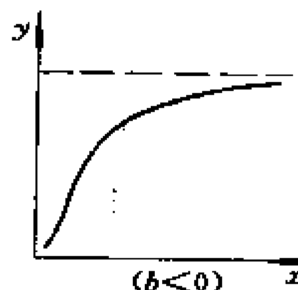
双曲线函数 $y=\frac{x}{ax+b}$的曲线

(d)



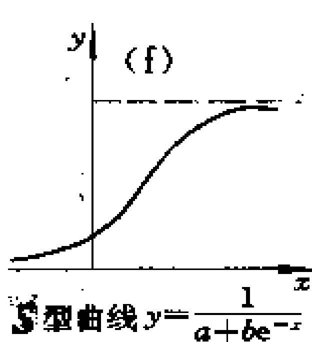
对数函数 $y=a+b\ln x$ 的曲线





指数函数

$y=ae^{\frac{b}{x}}$的曲线



S型曲线 $y=\frac{1}{a+be^{-x}}$

图 7-6

面对实验数据，如何选取适当的非线性模型，是建模的难点．除将各种模型逐一比较的方法外，最常用的方法主要有机理分析和散点图特征分析法．

例 7-7 为了检验 X 射线的杀菌作用，用 200 千伏的 X 射线来照射细菌，每次照射 6 分钟．照射次数记为 t，共照射 15 次，各次照射后所剩细菌数 y 如表 7-9 所示．

表 7-9 细菌数观测数据

序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
y	352	211	197	160	142	106	104	60	56	38	36	32	21	19	15

首先画出其散点图如图 7-7 所示．

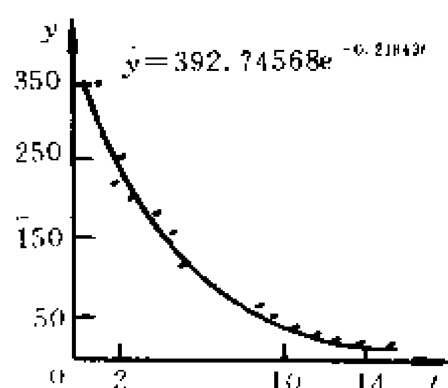


图 7-7 细菌数散点图与它的回归曲线

从散点图的特点分析，可供选择的非线性模型有：双曲函数 $y=\frac{t}{at+b}$，抛物线(幂函数的变形) $y=at^2+bt+c$ 和指数函数 $y=ae^{bt}$，其中最为接近的是指数函数．另外，由以往经验知，细菌数 y 的对数应是照射次数 t 的线性函数，即

$$\ln y=\beta_0+\beta_1 t,\quad y=ae^{\beta_1 t}$$

其中，$a=e^{\beta_0}$．令 $y'=\ln y$，则有

$$y'=\beta_0+\beta_1 t$$

采用最小二乘法，可得 β_0、β_1 的估计值，从而有

$$\hat{y}'=5.97316-0.21843t$$

或

$$\hat{y}=e^{5.97316}e^{-0.21843t}$$

即

$$\hat{y}=392.74568e^{-0.21843t}.$$

这种负指数模型常用于生物生长、细菌繁殖和经济指标的增长等.

例 7-8 炼钢厂出钢时所用盛钢水的钢包在使用过程中，钢渣及炉渣对包衬耐火材料的侵蚀，使其容积不断增长．经试验，钢包的容积(由于容积不便测量，故以钢包盛满时钢水重量来表示)与相应的使用次数(也称包龄)的数据如表 7-10 所示，希望能找出其定量关系

表 7-10 钢包容积观测数据

使用次数(x)	2	3	4	5	7	8
容积(y)	106.42	108.2	109.58	110.0	109.93	110.49
使用次数(x)	11	14	15	16	18	19
容积(y)	110.59	110.60	110.90	110.76	119.00	111.20

首先画出散点图(图 7-8)

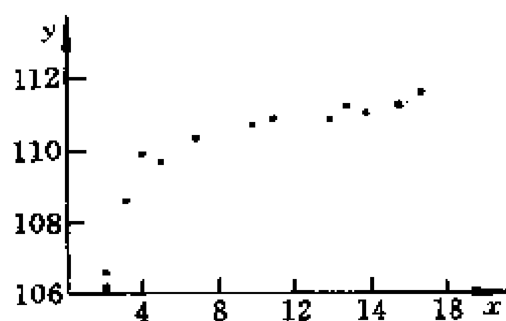


图 7-8 钢包容积散点图

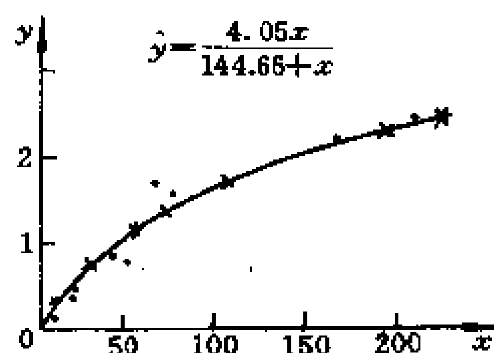


图 7-9

从图中可见，最初容积增加很快，以后逐渐减慢趋于稳定，这正是双曲线的特点．用双曲线

$$y=\frac{x}{ax+b},\quad \frac{1}{y}=\beta_0+\beta_1\frac{1}{x}$$

来表示容积 y 与使用次数 x 之间的关系，令

$$y'=\frac{1}{y},\quad x'=\frac{1}{x}$$

则有

$$y'=\beta_0+\beta_1 x'$$

利用最小二乘方法得到系数的估计值，则

$$\hat{y}'=0.008966+0.0008302x'$$

即

$$\hat{y}=\frac{x}{0.0008302+0.008966x}$$

历史上一个著名的例子说明了散点图和待选用的非线性曲线的特点可作为选用模型的依据．

在寻找有关细菌生长的合适数学描述时，Monod 比较了两个候选模型．问题的一般形式是：

$$\begin{cases}\dfrac{dx}{dt}=\mu(x,s)x\\[2mm] \dfrac{ds}{dt}=-\dfrac{1}{y}\dfrac{dx}{dt}\end{cases}$$

$$y(t)=x(t)+s(t)$$

其中，x 是作为人口密度度量的生物量浓度，s 是作为可利用食物度量的作用物浓度．当选择不同的 $\mu(x,s)$时，可获不同类型：

$$M_1:\mu(x,s)=\mu_x(1-\frac{x}{x_M})x_M=x_0+ys_0\text{(Logistic 模型)}$$

$$M_2:\mu(x,s)=\mu_M s/(k_M+s)$$

两个模型都是非线性的，对响应 $x(t)$进行分析，用数学方法发现它呈现 s 形，并且是非递减的．其中 M_1 在 $x=0$ 和 $x=x_M$(x 的最大值)中间有一拐点；M_2 的拐点依赖于参数值，大致位于曲线的上半部某处．通过观测实验数据的散点图，Monod 发现拐点接近于最大值 x_M，于是模型 M_1 被拒绝，而采用 M_2 作为模型．

作为选用"合适"的非线性模型的"笨"方法，是逐一比较各种

非线性模型，这就需要有一个评价标准．在一元线性回归中，用相关系数 r 检验回归方程．对于一元非线性回归问题，用类似于相关系数的相关指数来衡量所配曲线效果的好坏．相关指数定义为

$$R^2=1-\frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}$$

R^2 越接近于 1，所配曲线的效果越好．另外，残差平方和

$$Q=\sum(y-\hat{y})^2$$

与标准差

$$S=\sqrt{\frac{\sum(y-\hat{y})^2}{n-2}}$$

都可以用来衡量曲线的拟合好坏．Q(或 S)越小说明拟合得越好．

例 7-9　重新讨论例 7-7．y 与 t 的函数关系可能为指数模型 $y=\alpha e^{\beta t}$，也可能为幂函数模型 $y=\alpha t^{\beta}$.

若采用 $y=\alpha t^{\beta}$，令 $y''=\ln y$，$t''=\ln t$，得

$$y''=\ln\alpha+\beta t''=\beta_0+\beta_1 t''$$

计算得

$$\hat{y}''=6.415-1.177t''$$

即

$$\hat{y}=610.941t^{-1.177}$$

残差平方和为

$$Q=\sum(y-\hat{y})^2=77745.956$$

标准差为

$$S=\sqrt{\frac{Q}{n-2}}=77.333$$

相关指数为

$$R^2=1-\frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}=0.397$$

而采用指数函数 $y=\alpha e^{\beta t}$，相应的残差平方和、标准差、相关系数分别为 4015.315，17.575，0.969 均优于前者．所以指数模型优

于幂函数模型．

2. 一般非线性模型

并不是所有非线性模型都可通过变换化为线性模型．例如

$$y=e^{\beta_1 x}+e^{\beta_2 x},\quad y=\beta_0+\beta_1{}^{x},\quad y=(\beta_0+\beta_1 x^{\beta_2})^{1}$$

等，无论使用怎样的变换，都不能化为线性模型．这样的模型，称为本质非线性模型或纯非线性模型．

一般非线性模型的形式为

$$y=f(x,\theta)+\varepsilon$$

其中，f 是一般的函数，θ 是 p 维参数向量，ε 是一随机误差变量，$E(\varepsilon)=0$，$\mathrm{Var}(\varepsilon)=\sigma^2$.

求“最小二乘”拟合曲线，就是求 θ 的估计 $\hat{\theta}$，使

$$\min S(\theta)=\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}[y_i-f(x_i,\theta)]^2$$

再将 $\bar{\theta}$ 的值代入 $f(x,\theta)$，得到拟合曲线

$$\hat{y}=f(x,\theta)$$

问题转换为求解非线性规划问题．用于非线性规划的逐次线性化方法或变尺度方法等都可用来求解这里的“最小二乘”问题．通常采用高斯-牛顿方法求解．

设 $(x_i,y_i)\,i=1,2,\cdots,n$ 是观察数据．用 $f_i(\theta)$ 代替 $f(x_i,\theta)$，S 代替 $S(\theta)$.

在点 $\theta=\theta_0$ 处，将 $f_i(\theta)$ 泰勒展开，只取前两项

$$f(\theta)=f(\theta_0)+J(\theta_0)(\theta-\theta_0)$$

式中

$$\boldsymbol{f}(\theta)=(f_1(\theta),f_2(\theta),\cdots,f_n(\theta))^T$$

$J(\theta_0)$ 是 $n\times p$ 阶雅可比矩阵

$$J(\theta_0)=\begin{bmatrix}\dfrac{\partial f_1(\theta)}{\partial\theta_1} & \dfrac{\partial f_1(\theta)}{\partial\theta_2} & \cdots & \dfrac{\partial f_1(\theta)}{\partial\theta_p}\\ \cdots & \cdots & \cdots & \cdots\\ \dfrac{\partial f_n(\theta)}{\partial\theta_1} & \dfrac{\partial f_n(\theta)}{\partial\theta_2} & \cdots & \dfrac{\partial f_n(\theta)}{\partial\theta_p}\end{bmatrix}_{\theta=\theta_0}$$

用 $\boldsymbol{y}$ 记$(y_1,y_2,\cdots,y_n)^T$，则

$$S=\sum_{i=1}^{n}[y_i-f(x_i,\theta)]^2=[\boldsymbol{y}-\boldsymbol{f}(\theta)]^T[\boldsymbol{y}-\boldsymbol{f}(\theta)]$$
$$=[\boldsymbol{y}-\boldsymbol{f}(\theta_0)-J(\theta_0)(\theta-\theta_0)]^T[\boldsymbol{y}-\boldsymbol{f}(\theta_0)-J(\theta_0)(\theta-\theta_0)]$$
$$=[\boldsymbol{y}-\boldsymbol{f}(\theta_0)]^T[\boldsymbol{y}-\boldsymbol{f}(\theta_0)]-2[\boldsymbol{y}-\boldsymbol{f}(\theta_0)]^TJ(\theta_0)(\theta-\theta_0)$$
$$+(\theta-\theta_0)^TJ^T(\theta_0)J(\theta_0)(\theta-\theta_0)$$

记
$$\boldsymbol{g}(\theta)=\left(\frac{\partial S}{\partial\theta_1},\frac{\partial S}{\partial\theta_2},\cdots,\frac{\partial S}{\partial\theta_n}\right)^T$$

则有

$$\boldsymbol{g}(\theta)=-2J^T(\theta_0)[\boldsymbol{y}-\boldsymbol{f}(\theta_0)]+2J^T(\theta_0)J(\theta_0)(\theta-\theta_0)$$

令 $\boldsymbol{g}(\theta)=0$，有

$$J^T(\theta_0)J(\theta_0)(\theta-\theta_0)=J^T(\theta_0)[\boldsymbol{y}-\boldsymbol{f}(\theta_0)]$$
$$\theta=\theta_0+[J^T(\theta_0)J(\theta_0)]^{-1}J^T(\theta_0)[\boldsymbol{y}-\boldsymbol{f}(\theta_0)]$$

这样得到递推公式

$$\theta_{i+1}=\theta_i+[J^T(\theta_i)J(\theta_i)]^{-1}J^T(\theta_i)[\boldsymbol{y}-\boldsymbol{f}(\theta_i)]$$

考虑线性模型

$$\boldsymbol{y}=\boldsymbol{x}\theta+\boldsymbol{\varepsilon}$$

雅可比矩阵为

$$J(\theta)=\boldsymbol{x}$$

无论从哪一点 θ_0 开始，第一估计 θ_1 为

$$\theta_1=\theta_0+(\boldsymbol{x}^T\boldsymbol{x})^{-1}\boldsymbol{x}^T(\boldsymbol{y}-\boldsymbol{x}\theta_0)=\theta_0+(\boldsymbol{x}^T\boldsymbol{x})^{-1}\boldsymbol{x}^T\boldsymbol{y}-\theta_0$$
$$=(\boldsymbol{x}^T\boldsymbol{x})^{-1}\boldsymbol{x}^T\boldsymbol{y}.$$

继续下去，估计值不变，即第一步就可得到 θ 的估计值．这与线性回归中的结果一样．由此可见，求回归直线的方法是高斯-牛顿方法的特殊情形．

例 7-10 根据药理学的蛋白结合理论，蛋白的 Langmuir 型单分子层具有吸附功能，在恒温的条件下，每一克蛋白的药物吸附量 y 与血浆浓度 x 的关系为 $y=\dfrac{\beta x}{\alpha+x}$. 表 7-11 给出了 10 组观察值，求 α 与 β 的估计值．

表 7-11　药物吸附量与血浆浓度观测数据

序号	1	2	3	4	5	6	7	8	9	10
x	12.7	21.1	51.7	77.2	212.4	9.5	22.5	42.3	67.8	234.8
y	0.103	0.466	0.767	1.573	2.462	0.083	0.399	0.889	1.735	2.360

此题显然可采用变换，将模型转换为

$$\frac{1}{y}=\frac{\alpha}{\beta}\cdot\frac{1}{x}+\frac{1}{\beta}$$

令 $y'=\frac{1}{y}, x'=\frac{1}{x}, a=\frac{\alpha}{\beta}, b=\frac{1}{\beta}$，则得到线性模型

$$y'=ax'+b$$

在得到估计值 a 和 b 后，$\beta=\frac{1}{b}, \alpha=\frac{a}{b}$.

这里采用高斯-牛顿公式，直接求 α 和 β 的估计值．因为

$$\frac{\partial f}{\partial \alpha}=\frac{-\beta x}{(\alpha+x)^2}, \frac{\partial f}{\partial \beta}=\frac{x}{\alpha+x}$$

取 a,b 的初值为 $a_0=261.39, b_0=5.98$，则 J_0 与 $y-f_0$ 为

$$J_0=\begin{pmatrix} -0.0010 & 0.0463 \\ -0.0016 & 0.0747 \\ -0.0032 & 0.1651 \\ -0.0040 & 0.2280 \\ -0.0057 & 0.4483 \\ -0.0008 & 0.0351 \\ -0.0017 & 0.0793 \\ -0.0027 & 0.1793 \\ -0.0037 & 0.2060 \\ -0.0057 & 0.4732 \end{pmatrix}, \quad \boldsymbol{y}-\boldsymbol{f}_0=\begin{pmatrix} -0.1741 \\ 0.0190 \\ -0.2205 \\ 0.2095 \\ -0.2188 \\ -0.1267 \\ -0.0750 \\ 0.0561 \\ 0.5034 \\ -0.4698 \end{pmatrix}$$

计算得

$$(J_0^T J_0)^{-1} J_0^T(\boldsymbol{y}-\boldsymbol{f}_0)=(-188.62,-3.02)^T$$

代入递推公式有

$$a_1=261.39-188.62=72.77$$

$$b_1=5.98-3.02=2.96$$

若取临界值 $\delta=0.00005$，继续迭代则最终可得 α 与 β 的估计值为

$$a=144.66,\quad b=4.05$$

即拟合曲线为

$$\hat{y}=\frac{4.05x}{144.66+x}$$

其图象与表 7-8 的散点图如图 7-10 所示．

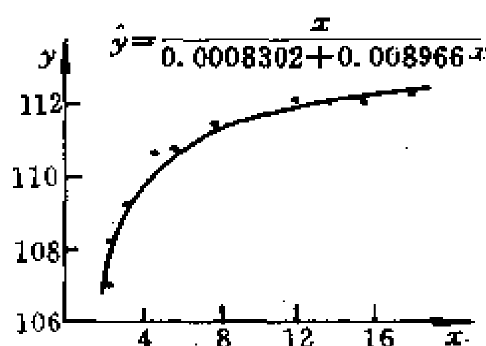


图 7-10　药物吸附量散点图与回归曲线

§7.6　某些常用非线性模型

这里主要讨论非线性增长模型．

一、S 形增长模型

增长模型在很多领域都有应用：生物学、生态学、化学以及政治和经济等等领域．增长模型一般是机理模型，而不是经验模型．

1. Logistri 模型

Logistri 模型的推导参见第五章 §5.2. 其形式为

$$\frac{\mathrm{d}\omega}{\mathrm{d}t}=\frac{k\omega(\alpha-\omega)}{\alpha}\qquad(k>0)$$

方程的解为

$$\omega=\alpha/(1+\beta e^{-kt})$$

由于$\lim\limits_{t\to\infty}\omega=\alpha$,这也是一种渐近回归模型.

它的变形有

$$\omega=\frac{\alpha}{1+\exp(\beta-rt)},\quad \omega=\frac{1}{\alpha+\beta\exp(-rt)}$$

$$\omega=\frac{1}{\alpha+\beta r^t},\quad \omega=\frac{\alpha}{1+\exp(\beta)\cdot rt}$$

$$\omega=\frac{1}{\alpha+\exp(\beta)\cdot rt},\quad \omega=\frac{\alpha}{1+\beta\exp(-rt)}$$

等等.有时它也被称为自动催化增长模型.

2. Gomperty 模型

若增长率为

$$\frac{d\omega}{dt}=k\omega\log(\alpha/\omega)$$

方程的解为

$$\omega=\alpha\exp[-\beta e^{-kt}]$$

它的曲线也类似于S形.其形状特点是,由$\frac{d^2\omega}{dt^2}=0$,拐点为$\omega=\frac{\alpha}{e}$ $=0.3682$,对应$t=\log\beta/k$.另外,由$\frac{d\omega/dt}{\omega}=k\log(\frac{\alpha}{\omega})=k(\log\alpha-\log\omega)$,隐含着相对增长率与$\log\omega$之间的线性关系.

注意到

$$\frac{d\omega/dt}{\omega}=k[\log\alpha-\log(\alpha\exp(-\beta e^{-kt}))]$$

$$=k[\log\alpha-\log\alpha+\beta e^{-kt}]=k\beta e^{-kt}$$

$$\therefore\qquad \log(\frac{d\omega/dt}{\omega})=\log(k\beta)-kt$$

这表明相对增长率的对数与时间t有线性关系.

该模型的变形为

$$\omega=\alpha\exp[-\exp(\beta-rx)]$$

$$\omega=\exp(\alpha-\beta r^{t})$$

Logistic 模型和 Gompertz 模型的图形都具有 S 形，曲线在某点后递增率由迅速增大而逐渐减小，并且趋于一个稳定值，即曲线存在拐点称为 S 形生长模型. 有很多函数都可作为 S 形生长模型，除上述两种外，还有

Richards 模型

$$\omega=\alpha/[1+\exp(\beta-rt)]^{1/\delta}$$

Morgan-Mercer-Flodin 模型

$$\omega=\frac{\beta r+\alpha t^{\delta}}{r+t^{\delta}}$$

Weibull 模型

$$\omega=\alpha-\beta\exp(-rt^{\delta})$$

S 形模型中的参数 α、β 和 r 有其自身的含义. 参数 α 与渐近性有关，对于大部分模型，渐近线是 $y=\alpha$（或 $\exp(\alpha)$，$\frac{1}{\alpha}$）；参数 β 与 y 轴上的"截距"有关，对于某些模型，截距正好是 β；参数 r 与响应变量从"初值"（由 β 的大小确定）改变到它的"终值"（由 α 的大小确定）的速度有关. 在四参数模型中，参数 δ 用来增加数据拟合模型的灵活性.

对于这些模型参数的估计，通常是从一组数据出发，假设非线性模型为

$$y=f(x,\theta)+\varepsilon \qquad \text{（加法误差）}$$

或

$$y=f(x,\theta)e^{\varepsilon} \qquad \text{（乘法误差）}$$

然后再分别求出残差平方和 $Q=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$，取 Q 值最小的模型作为应拟合的模型.

例 7-11 表 7-12 中列出四组观测数据，其散点图如图 7-10 所示，试研究其规律.

表 7-12　四组观测数据

数据组 Ⅰ		数据组 Ⅱ		数据组 Ⅲ		数据组 Ⅳ	
x	y	x	y	x	y	x	y
9	8.93	1	16.08	0	1.23	0.5	1.3
14	10.80	2	33.83	1	1.52	1.5	1.3
21	18.59	3	65.80	2	2.95	2.5	1.9
28	22.33	4	97.20	3	4.34	3.5	3.4
42	39.35	5	191.55	4	5.26	4.5	5.3
57	56.11	6	326.20	5	5.84	5.5	7.1
63	61.73	7	386.87	6	6.21	6.7	10.6
70	64.62	8	520.53	8	6.50	7.5	16.0
79	67.08	9	590.03	10	6.83	8.5	16.4
		10	651.92			9.5	18.3
		11	724.93			10.5	20.9
		12	699.56			11.5	20.5
		13	689.96			12.5	21.3
		14	637.56			13.5	21.2
		15	717.41			14.5	20.9

从图可见，它们都是 S 形生长模型（具有渐近性和拐点），对这四组数据分别按加法误差和乘法误差，用五种生长模型进行参数估计，并计算残差平方和 Q 与方差 $\hat{\sigma}^2=Q/n-p$，如表 7-13 所示.

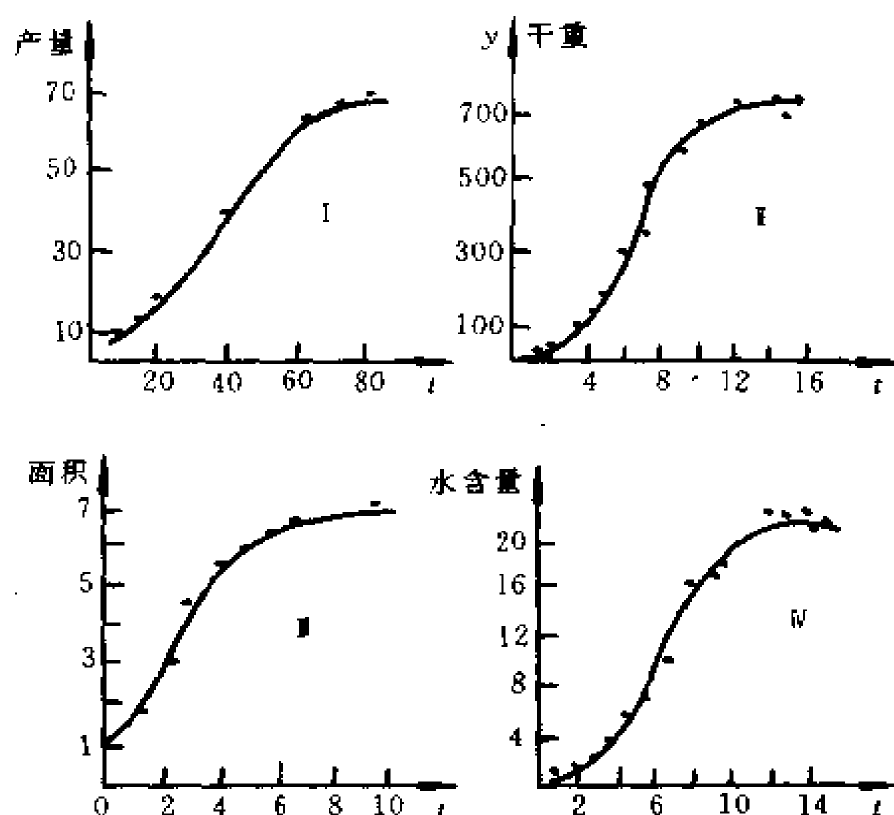


图 7-10

从表中容易看出每一组数据用哪一种模型拟合较好.

3. 渐近回归模型

形如 $y=\alpha-\beta r^{x}$ 的渐近回归模型,广泛地用于农业、生物学和工程学.它最早也是最经常的一种应用是在施肥试验方面,x 表示肥料的使用量,y 表示作物产量,此模型常称为 Mitschertich 律.在渔场研究中也常常用此模型描述鱼的长度与年龄的关系,称为 VonBertalanffy 生长曲线.当 x 增加时,曲线接近于某一渐近线,但与 S 形生长模型不同的是,该类模型没有拐点.

一种渐近回归模型是单分子增长模型,它的形式为

$$y=\alpha(1-\beta e^{-kx})$$

表 7-13　生长模型的参数估计、加法误差与乘法误差

数据组	参数	三参数模型				四参数模型					
		戈配兹模型		若吉斯蒂克模型		理查兹模型		摩根-默西-弗洛门模型		威布尔模型	
		加法误差	乘法误差	加法误差	乘法误差	加法误差	乘法误差	加法误差	乘法误差	加法误差	乘法误差
Ⅰ	α	82.83	96.94	72.46	73.20	69.62	72.53	80.96	93.26	69.96	73.48
	β	1.224	1.159	2.618	2.591	4.255	2.773	8.895	7.328	61.68	66.41
	γ	0.037	0.030	0.067	0.066	0.089	0.068	49577	8773	0.000100	0.000261
	δ					1.724	1.072	2.828	2.298	2.378	2.106
	$\hat{\sigma}^2$	3.63	0.0047	1.34	0.0032	1.21	0.0038	2.71	0.0042	1.68	0.0038
Ⅱ	α	723.1	903.1	702.9	703.3	699.6	702.5	723.9	827.8	695.0	708.9
	β	2.500	1.682	4.443	4.395	5.277	4.449	33.35	15.94	673.5	693.9
	γ	0.450	0.254	0.689	0.680	0.760	0.686	6266	550	0.00152	0.00293
	δ					1.279	1.013	4.641	3.178	3.267	2.891
	$\hat{\sigma}^2$	1134	0.0215	744	0.0047	799	0.0051	1015	0.0130	712	0.0083
Ⅲ	α	6.925	7.585	6.687	6.894	6.684	6.497	6.986	6.836	6.656	6.478
	β	0.768	0.666	1.745	1.664	1.780	4.832	1.181	1.216	5.549	5.280
	γ	0.493	0.366	0.755	0.669	0.759	1.293	12.96	15.53	0.118	0.0826
	δ					1.017	2.772	2.475	2.707	1.763	2.138
	$\hat{\sigma}^2$	0.0619	0.0177	0.0353	0.0104	0.0424	0.0092	0.0048	0.0004	0.0268	0.0026
Ⅳ	α	22.51	36.20	21.51	23.34	21.20	20.78	22.08	23.95	21.10	21.22
	β	2.106	1.391	3.957	3.398	5.691	9.939	1.653	1.264	19.81	19.99
	γ	0.388	0.170	0.622	0.484	0.777	1.219	5586	854.8	0.00177	0.00218
	δ					1.619	3.111	4.560	3.470	3.180	3.062
	$\hat{\sigma}^2$	1.049	0.0440	0.518	0.0193	0.502	0.0128	0.579	0.0054	0.495	0.0040

它是方程

$$\frac{\mathrm{d}y}{\mathrm{d}t}=k(\alpha-\omega)$$

的解.

其他类型的渐近回归模型有

$$y=\alpha\{1-\exp[-(x+\beta)r]\}$$

$$y=\alpha-\exp[-(\beta+rx)]$$

$$y=\alpha-\exp(-\beta)r^x$$

$$y=\frac{1}{\alpha}-\beta r^x \tag{1}$$

$$y=\exp(\alpha)-\beta r^x$$

注意到参数 r 由 $\exp(-rx)$ 变换到 r^x,参数 β 由 $\exp(-\beta)$ 变换为 β,参数 α 由 α 变换到 $\exp(\alpha)$ 都将导致线性程度的改善,而 α 变为 $\frac{1}{\alpha}$ 则更为理想,因此(1)式模型应用更为广泛.

4. 产量-密度模型

在农业生产中,农作物产量和密度之间有密切的关系.一般在实际中会出现两种基本情况,即所谓"渐近线"和"抛物线"产量-密度模型.

用 x 表示单位面积播种数,y 表示每棵作物的产量,那么 $\omega=xy$ 是单位面积的总产量.如果 ω 随着单位面积种植数的增加逐步上升到一稳定值,就称其服从渐近线产量-密度关系;如果 ω 随着单位面积播种数的增加逐步上升到一个最大值,然后逐步下降,就说它服从抛物线产量-密度关系.

有很多函数可以用来描述产量-密度关系.下面是三个参数个数为 3 的模型.

1° Bleasdale(巴列斯尔特)和 Neldex(乃尔德)在 1960 年给出的模型

$$y=(\alpha+\beta x)^{-\frac{1}{\theta}}$$

2° Hallidag(哈利德)在 1960 年给出的模型

$$y=(\alpha+\beta x+v x^2)^{-1}$$

3° Farazdaghi(兰那达黑)和 Haris(哈雷斯)在 1968 年给出的模型.

$$y=(\alpha+\beta x^1)^{-1}$$

当 $\theta=\varphi=1, v=0$ 时,以上三个模型都化为

$$y=(\alpha+\beta x)^{-1}$$

这是一个典型的渐近线产量-密度模型,是一种双曲型函数,Shinozaki 和 Kira(1956)称为渐近模型,Mead(1979)称之为倒数模型.

渐近模型的参数有其物理意义.当密度 x 趋向于零时,每株植物产量 y 趋向于$\frac{1}{\alpha}$,因此 α(或$\frac{1}{\alpha}$)可以认为是,在没有因环境资源的竞争而产生对物种的抑制作用时,物种的"遗传潜力"的一种度量.在较高的植物密度上,当 $x\to\infty$时

$$\omega=xy=\frac{x}{\alpha+\beta x}\to\frac{1}{\beta}$$

因此 β(或$\frac{1}{\beta}$)可以认为是"环境潜力"的一种度量.

当考虑如何选取最佳种植密度时,以模型②为例,

$$\omega=xy=\frac{x}{\alpha+\beta x+v x^2}$$

求得极值点为 $$x=\sqrt{\frac{\alpha}{v}},$$

即是说,当求得 α,v 的估计 $\hat{\alpha},\hat{v}$ 后,取

$$x=\sqrt{\frac{\hat{\alpha}}{\hat{v}}}$$

产量最高.

在实际处理时,一般对上述三个模型的等式两边取对数,化为

$$\ln y=-\frac{1}{\theta}\ln(\alpha+\beta x)$$

$$\ln y=-\ln(\alpha+\beta x+v x^2)$$

$$\ln y = -\ln(\alpha + \beta x^{\theta})$$

然后求使

$$\min \sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} (\ln y - \ln \hat{y})^2$$

的估计参数.

例 7-12 表 7-12 列出了南澳大利亚四个产地(MG:蒙特冈比尔;U:乌拉依杜拉;PL:普尔农兰丁;V:维尔金尼亚)洋葱的数据,产量 y(克/棵)作为密度 x(棵/m^2)的函数. 其散点图如图 7-11 所示.

表 7-14 四产地产量与密度观测数据

深黄色的标准西班牙品种				白色的标准西班牙品种			
(MG)		(U)		(PL)		(V)	
X	Y	X	Y	X	Y	X	Y
95.47	71.28	98.91	65.67	104.51	81.71	89.45	79.94
98.05	56.61	103.44	67.19	105.68	76.44	90.93	79.13
98.42	75.09	105.05	54.01	108.03	87.10	92.91	70.93
102.48	65.26	111.19	60.92	117.82	84.54	101.81	60.99
105.80	64.48	113.78	53.48	127.21	69.09	103.78	74.09
106.53	61.84	119.92	61.62	134.26	64.40	115.15	49.45
108.75	65.19	120.89[a]	26.32	137.39	66.81	123.06	56.65
115.38	57.10	126.71	61.21	151.87	63.01	144.31	47.84
150.77	52.68	138.99	41.67	163.61	55.45	155.68	40.03
152.24	47.01	146.75	45.26	166.35	62.54	158.15	38.70
155.19	44.28	160.97	46.45	184.75	54.68	180.39[a]	28.96

注:表中凡数上标有[a]的点是可疑的,在计算中删去了这些点.

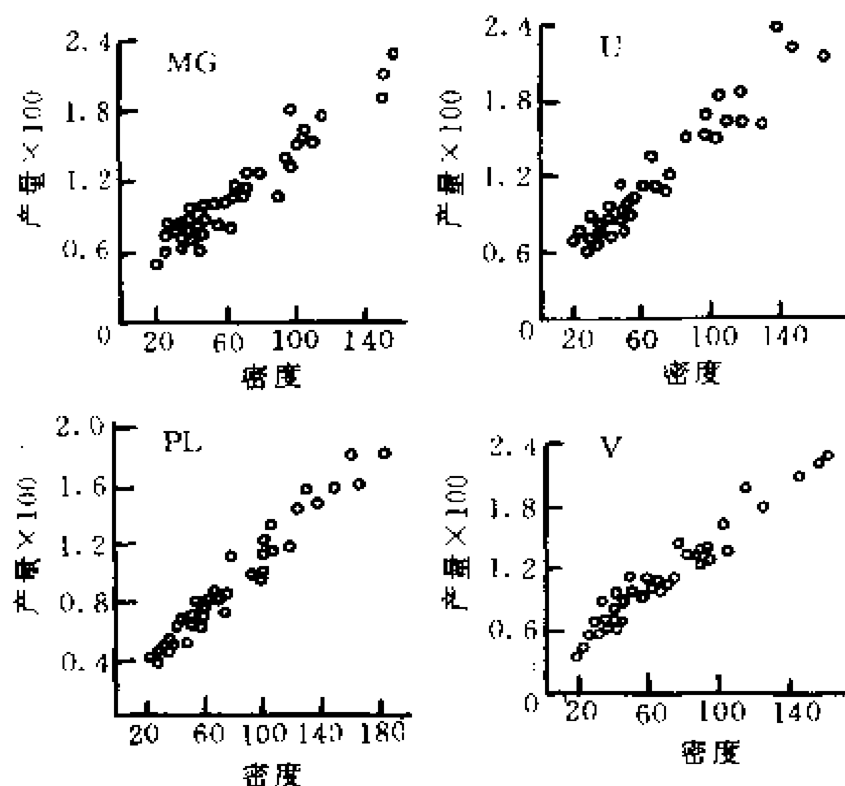


图 7-11 四产地产量散点图

假设表 7-8 的数据 var(lny)是常数，用①②③三种模型求出参数的最小二乘估计和残差平方和，如表 7-15 所示，从中可见，每个产地用什么模型拟合最好．

表 7-15 四产地三种回归模型参数表

产地	参数	(1)	(2)	(3)
MG	α	0.03267	0.004524	0.005043
	β	0.3581×10^{-3}	0.08113×10^{-3}	0.02869×10^{-3}
	θ, ν 或 φ	0.6346	0.1976×10^{-6}	1.263
	$\hat{\sigma}^2=Q/(n-p)$	0.01229	0.1231	0.01226

续表

产地	参数	(1)	(2)	(3)
U	α	0.02019	0.004220	0.004859
	β	0.3222×10^{-3}	0.1012×10^{-3}	0.03877×10^{-3}
	θ,υ 或 φ	0.7162	0.1933×10^{-6}	1.235
	$\hat{\sigma}^2=Q/(n-p)$	0.02303	0.02308	0.02295
PL	α	0.003842	0.002054	0.002294
	β	0.1302×10^{-3}	0.08571×10^{-3}	0.06032×10^{-3}
	θ,υ 或 φ	0.9055	0.03808×10^{-6}	1.080
	$\hat{\sigma}^2=Q/(n-p)$	0.007265	0.007287	0.007253
V	α	0.001803	0.002084	0.001648
	β	0.1418×10^{-3}	0.1311×10^{-3}	0.1581×10^{-3}
	θ,υ 或 φ	1.000	0.07796×10^{-6}	0.9791
	$\hat{\sigma}^2=Q/(n-p)$	0.01568	0.01558	0.01568

本章讨论的是数学建模中一个十分重要而又十分困难的问题——结构特征化问题．

数学建模的目的，就是在于给出实际问题的一个数学结构．面对一组观测数据，建模的首要问题是如何选定一个合适的结构（模型函数），然后对结构的参数进行估计，其中参数估计有很多成熟的方法，而模型函数的选择则具有更大的灵活性，需要更多的经验与技巧．当然这两个步骤不是绝然分开的，参数估计的结果常常被用来选择模型函数．在§3中已看到，参数估计对选择基函数的个数是有用的．

本章讨论了在一类候选模型函数中寻找最佳者的方法，以及选择基函数和作为常微分方程反问题的方法来寻找适当的模型函数．实际上，如果我们能了解问题的实质、利用机理分析获得结

构，然后进行参数估计，无疑是最有把握的途径．然而当我们面对一组观测数据，而没有任何先验知识的情况下，我们能做的仅仅只是在一堆候选集中寻找一个较为合适的，显然有很多模型接近于这组观测数据，因此，对于有限数据集合，结构特征化问题没有唯一解．充分利用、挖掘观测数据这一资源对选择结构至关重要，例如是否具有极值，是否具有拐点等等．近年来，选择“合适”的结构这一问题，不仅从统计角度在深入研究，它也被认为是决策问题在数学建模中的应用，这导致了建模支持系统的迅速发展．

习　题

1．据观察，个子高的人一般腿都长，今从16名成年女子测得数据如表7-16，希望从中得到身高 x 与下体长 y 之间的回归关系．

表 7-16　身高与下身长观测数据　　单位：厘米

身高(x)	143	145	146	147	149	150	153	154	155	156	157	158	159	160	162	164
下体长(y)	88	85	88	91	92	93	93	95	96	98	97	96	98	99	100	102

2．某长途运输公司在同一类型的卡车中，对行驶公里数 y 和行驶天数 x 进行统计．得数据如下

表 7-17　行驶里程与天数观测数据

天数(x)	3.5	1.0	4.0	2.0	1.0	3.0	4.5	1.5	3.0	5.0
公里数(y)	825	215	1070	550	480	920	1350	325	670	1215

3．为了提高某丝织品的质量，通过控制上机张力来控制织缩率（成品长与原料丝长之比），进而减少断头率．进行了15次试验，数据如表7-18试建立织缩率与断头率的表达式．

4．某厂表面处理车间试验将铬后污水同电解污泥混合，使之生成无毒液体，效果很好．但实际排出污水的浓度不全相同，而且一定浓度的定量铬后污水只有同定量的电解污泥混合后，才能反应完全．现通过试验，找出铬后污水用量与电解污泥用量之比对于铬后污水浓度之间的关系．试验数据如表7-19所示．

表 7-18　织缩率与断头率观测数据

序号	1	2	3	4	5	6	7
织缩率(x)	4.20	4.06	3.80	3.60	3.40	3.20	3.00
断头率 y(根/台·时)	0.086	0.090	0.120	0.130	0.150	0.170	0.190

序号	8	9	10	11	12	13	14	15
织缩率(x)	2.80	2.60	2.40	2.20	2.00	1.80	1.60	1.40
断头率 y(根/台·时)	0.090	0.220	0.240	0.350	0.440	0.620	0.940	1.620

表 7-19　铬后污水用量与电解污泥用量试验数据

序号	1	2	3	4	5	6	7	8	9	10	11
铬后污水浓度 $x(g/l)$	3	5	10	30	40	50	60	80	100	120	160
$\frac{\text{铬后污水用量}(ml)}{\text{电解污泥用量}(ml)}y$	310	200	100	49	40	32	28	23	16	14	10

5. 由实践经验知，7 月份平均气温是影响第二代棉铃虫历期(完成某一虫期发育所需的天数)的主要因素．为了建立预报方程，我们收集了 7 年的统计数据如表 7-20 所示．

表 7-20　平均气温与棉铃虫历期观测数据

年序	1	2	3	4	5	6	7
七月份平均气温 x(℃)	27.2	25.7	25.3	25.7	29.3	27.2	26.5
历期 y(天)	33	40	41	36	33	34	37

6. 某丝织厂为了掌握一种新型织机的性能，考察了织造工序经轴嵌边裂缝病疵(经丝嵌入经轴两侧而塌入裂缝内)的庄数 y 与上道工序和织造工序温度差异 x 的关系，经 26 次试验，得到如下数据(表 7-21)．

7. 对热敏电阻器的电阻 y 与温度 x 作了试验，数据如表 7-22 所示，试研究其关系．

表 7-21　庄数与温度差异的试验数据

序号	1	2	3	4	5	6	7	8	9	10	11	12	13
x	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8
y	1	1	1	3	2	4	3	3	5	4	8	8	1
序号	14	15	16	17	18	19	20	21	22	23	24	25	26
x	8.5	9	9.5	10	10.5	11	11.5	12	12.5	13	13.5	14	14.5
y	10	11	19	18	18	24	24	25	28	31	29	30	38

表 7-22　电阻与温度试验数据

序号	1	2	3	4	5	6	7	8
温度 x	50	55	60	65	70	75	80	85
电阻 y	34780	28610	23650	19630	16370	13720	11540	9744
序号	9	10	11	12	13	14	15	16
温度 x	90	95	100	105	110	115	120	125
电阻 y	8266	7030	6005	5147	4427	3820	3307	2872

表 7-23　长度与年龄的观测数据

序号	1	2	3	4	5	6	7	8	9
x	1	1.5	1.5	1.5	2.5	4.0	5.0	5.0	7.0
y	1.80	1.85	1.87	1.77	2.02	2.27	2.15	2.26	2.35
序号	10	11	12	13	14	15	16	17	18
x	8.0	8.5	9.0	9.5	9.5	10.0	12.0	12.0	13.0
y	2.47	2.19	2.26	2.40	2.39	2.41	2.50	2.32	2.43
序号	1	2	3	4	5	6	7	8	9
x	13.0	14.5	15.5	15.5	16.5	17.0	22.5	29.0	31.5
y	2.47	2.56	2.65	2.47	2.64	2.56	2.70	2.72	2.57

8. 海生类动物儒艮的长度(y)对年龄(x)的关系，有如下观测数据(表 7-23)试确定之.

9. 研究每株嫩枝日产树叶数(y)对光照量(x)的关系，以 20℃时每平方米瓦特为单位，观测数据如表 7-24 所示，试确定之.

表 7-24 日产树叶数与光照量观测数据

序号	1	2	3	4	5	6
x	12	23	40	92	156	215
y	0.094	0.119	0.199	0.260	0.309	0.331

10. 研究小麦产量(y)对肥料水平(x)的关系，有如下观测数据(表 7-25)试确定之.

表 7-25 小麦产量与肥料水平观测数据

序号	1	2	3	4	5
x	0	10	20	30	40
y	26.2	30.4	36.3	37.8	38.6

11. 化学反应中，N_2O_5 分解量(y)对时间(x)有关系，观测数据如表 7-26 所示，试研究之.

表 7-26 N_2O_5 分解量与时间观测数据

序号	1	2	3	4	5	6
x	2	3	4	5	6	7
y	18.6	22.6	25.1	27.2	29.1	30.1

12. 小麦产量(y)对石灰使用率(x)有如下观测数据(表 7-27)，试研究之.

13. 土豆产量(y)对 P_2O_5 使用率(x)有如下观测数据(表 7-28)，试研究之.

表 7-27　小麦产量与石灰使用率观测数据

序号	1	2	3	4	5
x	0	1	2	3	4
y	44.4	54.6	63.8	65.7	68.9

表 7-28　土豆产量与 P_2O_5 使用率观测数据

序号	1	2	3	4	5
x	4	1	2	3	4
y	232.65	369.08	455.60	491.45	511.50

14. 橡胶树围长(y)对肥料使用率(x)有如下观测数据(表 7-29),试研究之.

表 7-29　橡胶树围长与肥料使用率观测数据

序号	1	2	3	4	5
x	0	1	3	5	7
y	20.518	21.138	21.734	22.218	22.286

15. 试用增长模型研究例 7-5.

第八章　时序分析法

连续动态系统的数学建模主要有两种方法：一种是从基本物理定律以及系统（设备）的结构数据推导出模型，称为机理分析法；另一种是从系统的运行和试验数据建立系统的模型（模型结构和参数），这种方法称之为系统辨识．

关于系统辨识的定义及辨识（Identification）的译法，目前尚不统一．现在广泛采用的是 Zadeh 的一个定义：系统辨识是在输入输出的基础之上，从一类系统中确定一个与所测系统等价的系统．

从长远看，解析法是建立模型的重要手段，但这种方法只能用于建立比较简单的系统（白箱）模型；对于大多数系统，过程是很复杂的（黑箱和灰箱），以致用解析模型很难准确描述．

经过客观实践检验的基本物理定律，最初都是经过反复实验和观测而建立的．从这一意义上讲，用系统辨识方法建立的数学模型，则是分析法建模的基础．由于所要研究的系统，很多是非常复杂的，而且千变万化，因此用系统辨识方法建立系统的数学模型，将处于重要的基础地位．在实际应用中，常将机理分析和系统辨识两种方法结合起来，尽量利用对物理过程的认识，将系统的模型结构分成已知的和未知的两部分，然后利用实测数据，将未知部分辨识出来．

系统的数学模型用数据表格或图形表示，称为非参数模型，例如系统的阶跃响应，脉冲响应和频率响应的记录图形（如 Bode 图）等．参数模型在时间域中主要有差分方程和微分方程，在频率域中，主要有传递函数法．

传递函数是为系统输出与输出的拉氏变换之比，可描述简单的线性动态系统，属于古典控制理论范畴．这里主要讨论的是现

代控制理论范畴内的差分方程及相应的时间序列方法．

从方法上看，时序分析法和回归分析法都使用统计方法，但由于时序分析法处理的是动态的相关数据，因此又称为过程统计，而回归分析法处理的是静态的独立数据，通常称为数理统计．

§8.1 预备知识

一、随机过程

对生产和科学研究中的某一随机变量，例如机床振动的振幅，进行连续观察，观测记录如图 8-1 所示．

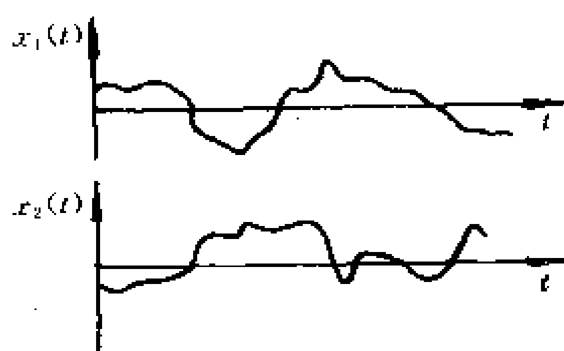


图 8-1 观测数据

这里随机变量 x，连续地依赖于时间 t. 称表示随机现象的时间历程为样本函数，称在有限时间区间上观察所得的结果为样本记录，如 $x_1(t)$，$x_2(t)$等．随机现象可能产生的全部样本函数的集合 $\{x(t), t\in T\}$称为随机过程．每条试验曲线即每一个样本记录可以理解为随机过程的一个物理实现．

随机过程可分为平稳与非平稳过程两种．若一个随机过程的统计特性不随时间而变化则称之为平稳过程．如图 8-2 所示的振动过程，当被测时间变化以后，从 t_1 到 t_2 这段时间的随机振动的统计特性，与$(t_1+\tau)$到$(t_2+\tau)$这段时间的统计特性差别不大，即把随机过程在时间上往后推移 τ，它们的统计特性并不改变．换

句话说，就是某一时刻的统计特性过一段时间后仍然适用．

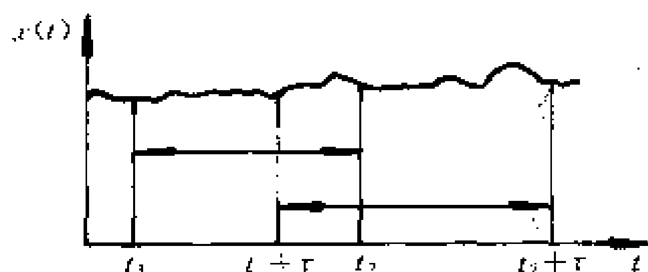


图 8-2　平稳过程图示

平稳过程通常是当一个稳定的物理系统达到稳态所产生的过程，这种状态有时也称为"统计平衡"状态．这要求 $x(t)$ 的高阶矩甚至一切有限维分布随时间的推移是不变的，即随机过程的概率结构对时间原点的移动保持不变，这种性质称为严平稳或完全平稳．

完全平稳的要求是苛刻的，实际中研究的是一种宽平稳（也称为广义平稳或弱平稳）．设随机过程 $x_T=\{x(t),t\in T\}$ 存在有限二阶矩 $E|x(t)|^2<\infty$，T 为实数集或其子集．若 x_T 满足 $Ex(t)=m$（常数），$E\{x(t)\overline{x(s)}\}=r(t-s)$，则称 x_T 为平稳随机过程．当 T 为整数集或子集时，也称为平稳时间序列．在工程和许多物理现象中经常遇到这类过程．特别是对于一个正态分布的过程，如果它是广义平稳的，则该过程也是严平稳的．

一般在求均值与均方差等数字特征时，为了确定概率密度函数，必须记录许多次采样函数（样本函数）．根据采样函数的集合所取得的平均值称为集合平均值．

在概率论中，我们已经知道了大数定理，即对独立同分布随机变量序列 $\{x_i,i\geqslant 1\}$，若存在均值 $Ex_i=m$，则

$$\lim_{N\to\infty} p\left\{\left|\frac{1}{N}\sum_{i=1}^{N}x_i-m\right|>\varepsilon\right\}=0 \tag{1}$$

从随机过程的观点看 $\{x_i,i\geqslant 1\}$ 也是一个随机过程，$\frac{1}{N}\sum_{i=1}^{N}x_i$ 是对

随机过程的样本按时间取平均，它随不同样本取不同数值，也是一个随机变量．而 $m=Ex_t$ 是随机过程的均值，即在某一时刻随机过程的现实取值的统计平均．所以(1)式表明随着时间的增长，随机过程样本按时间的平均值以越来越大的概率无限接近于随机过程的统计平均．即是说，对于随机过程，只要进行观察的时间足够长，它的样本都能“遍历”各种可能的状态，因而一个样本按时间的平均就可以近似地代替它在固定时刻取值的统计平均．随机过程的这个性质称为遍历性，当然这只有在不同时刻的统计平均是相等的这一前提下才有可能．所以对平稳过程考虑它的遍历性．由于在事实上常常不可能得到大量的记录，但在遍历性(也称为各态历经性)的假设下，可以认为单个采样在描述随机过程的所有特征时是有代表性的，除非有特殊且足够的证据来否定这个假设．

研究随机过程的重要工具是相关函数和功率谱密度(谱密度)．

二、相关函数

如果有两个时间函数 $x(t)$ 和 $y(t)$，其中一个函数在任一时刻的值总以某种方式依赖于另一个函数的值，则称这两个时间函数(或信号)是相关的．例如有一个信号 $x(t)$，若在 t 时刻的值总是在某一定程度上影响着时间间隔 τ 以后的值，即是说，$x(t+\tau)$ 与 $x(t)$ 是相关的．同一个信号的未来值与现在值之间的依赖关系可以采用“自相关函数”来度量，认为

$$R_{xx}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t)x(t+\tau)\mathrm{d}t$$

其中 τ 称为迟后．

一个信号的自相关函数具有下述性质：

1° $R_{xx}(0)=E(x^2)=\sigma^2$，$\sigma^2$ 为信号的均方值；

2° $\forall\ \tau$，$R_{xx}(\tau)\leqslant R_{xx}(0)$；

3° $R_{xx}(\tau)$ 是 τ 的偶函数，$R_{xx}(\tau)=R_{xx}(-\tau)$．

两个信号 $x(t)$和 $y(t)$之间的相关性可用互相关函数 $R_{xy}(\tau)$ 度量．

$$R_{xy}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t)y(t+\tau)\mathrm{d}t$$

三、功率谱密度

对于随机过程，如果它的时间历程明显是非周期函数，由傅里叶分析可知，这种随机过程有连续的无穷频率成分．对于一般物理过程，各个频率成分有它对应的功率，这种每个频率所含功率与频率的关系称为功率谱密度或谱密度．

1．自动率谱密度 $S_{xx}(\tau)$

根据维纳-辛钦（Wiener-Khintchine）定理

$$S_{xx}(\omega)=\int_{-\infty}^{+\infty}R_{xx}(\tau)\mathrm{e}^{-j\omega\tau}\mathrm{d}\tau$$

$$R_{xx}(\tau)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}S_{xx}(\omega)\mathrm{e}^{i\omega\tau}\mathrm{d}\omega$$

注意到 $R_{xx}(0)=E(x^2)$，得

$$\lim_{T\to\infty}\frac{1}{T}\int_{0}^{T}x^2(t)\mathrm{d}t=\int_{-\infty}^{+\infty}S_{xx}(\omega)\mathrm{d}\omega$$

若将上式中 $x(t)$理解为在一欧姆电阻上流过的电流，则上式左边就代表在 T 时间内消耗在一欧姆电阻中的平均功率．事实上质点运动的最大动能，弹簧势能的最大值都与振幅的平方成比例．因此功率谱密度包含每单位频宽所具有的功率的物理意义．自功率谱密度表征着功率按频率分布情况．

通常，把均值为零而谱密度为非零常数的平稳随机过程为白噪声（white voice）．其各出于白光具有均匀光谱的缘故．白噪声经傅氏分析，它在所有频率下面都具有恒定的幅值．白噪声变化速度极快．任一时刻的值与过去时刻的值毫无关系，其自相关函数为 δ 函数．

$$R_{xx}(\tau)=\begin{cases}\sigma^2, & \tau=0\\ 0, & \tau\neq 0\end{cases}$$

2. 互功率谱密度 $S_{xy}(\omega)$

$$S_{xy}(\omega)=\int_{-\infty}^{+\infty}R_{xy}(\tau)\mathrm{e}^{-j\omega\tau}\mathrm{d}\tau$$

$$R_{xy}(\tau)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}S_{xy}(\omega)\mathrm{e}^{j\omega\tau}\mathrm{d}\omega$$

四、采样与 Z 变换

在实际问题中，尽管测量记录常常是时间的连续函数，为了数字计算处理上的方便，往往只按一定的时间间隔取值．怎样合理地选择采样间隔，使得采样所得的序列能如实地代表原来的时间序列，这是数据采集中首先遇到的问题．申农(Shannon)指出：为了使采样信号恢复到原信号，其采样频率 ω_s 必须大于或等于原信号 $f(t)$ 频谱中的最高频率 ω_0 的两倍，即 $\omega_s\geqslant 2\omega_0$ 或采样周期 T 满足 $T\leqslant\frac{\pi}{\omega_0}$.

为了对采样得到的离散系统进行研究，引入 Z 变换．

经过采样，信号 $y(t)$ 成为离散脉冲序列

$$\begin{aligned}y^*(t)&=y(0)\delta(t)+y(T)\delta(t-T)+y(2T)\delta(t-2T)+\cdots\\&=\sum_{k=0}^{\infty}y(kT)\delta(t-kT)\end{aligned}$$

对上式进行拉普拉斯变换，记 $\mathscr{L}[y^*(t)]=y^*(s)$，称为 $y(t)$ 的采样拉普拉斯变换：

$$\begin{aligned}y^*(s)&=y(0)1+y(T)\mathrm{e}^{-Ts}+y(2T)\mathrm{e}^{-2Ts}+\cdots\\&=\sum_{k=0}^{\infty}y(kT)\mathrm{e}^{-kTs}\end{aligned}$$

其中利用到拉氏变换的延时定理

$$\mathscr{L}[f(t-kT)]=\mathrm{e}^{-kTs}F(s),$$

而得到 $\mathscr{L}[\delta(t-kT)]=\mathrm{e}^{-kTs}$.

令 $z=e^{Ts}$，于是

$$y^*(s)=\sum_{k=0}^{\infty}y(kT)Z^{-k}$$

称之为函数 $y(t)$的 Z 变换，记作 $Z[y(t)]$或 $Y(Z)$.

如同拉氏变换一样，Z 变换也是一种线性变换，其主要性质如下：

1° 线性性质：

若 $f_1(t)$、$f_2(t)$的 Z 变换分别为 $F_1(Z)$和 $F_2(Z)$，则

$$Z[f_1(t)+f_2(t)]=F_1(Z)+F_2(Z)$$

2° 延迟性质

若 $Z[f(t)]=F(Z)$，则 $Z[f(t-mT)]=Z^{-m}F(Z)$，其中 T 为采样周期.

3° 超前性质

$Z[f(t+mT)]=Z^mF(Z)-Z^m\sum_{k=0}^{m-1}f(kT)Z^{-k}$，如果 $f(0)=f(T)=\cdots=f[(m-1)T]=0$，则 $Z[f(t+mT)]=Z^mF(Z)$.

4° 初值定理

$$\lim_{t\to 0}f(t)=\lim_{k\to 0}f(kT)=\lim_{Z\to\infty}F(Z)$$

5° 终值定理

$\lim_{t\to\infty}f(t)=\lim_{k\to\infty}f(kT)=\lim_{Z\to 1}[(Z-1)F(Z)]$，此处当然要求被变换的时间函数必须有固定的有限终值.

在讨论从 Z 变换式 $y(Z)$求其原函数 $y(t)$时，应当注意到 Z 变换的一个重要性质，即 $y(Z)$与原函数并非一一对应. 因此利用 Z 变换表来查某一函数 $y(Z)$的原函数 $y(t)$，只是许多可能的答案之一，只是在 $t=kT(k=0,1,2,\cdots)$的诸时刻上，它的值 $y(kT)$才是确定的，至于在其他各时刻上原函数究竟取什么值，Z 变换表是不能确切给出的. 当然，依据申农的采样定理，如果采样周期足够短，由离散信号是可以恢复原信号的. 在许多场合下，人们只要把诸离散时刻的原函数值简单地用折线连接起来，也就足够了.

从 Z 变换的象函数求其对应的原函数，主要有三种方法：

1° 查表法

首先将已给的 Z 变换式化为函数 Z 变换对照表内所列的基本形式，然后查表.

例如，设 $Y(Z)=\frac{2.62Z^3-2.20Z^2+1.08Z}{Z^3-1.55Z^2+1.24Z-0.48}$，根据表中基本格式变换为

$$Y(Z)=Z\left[\frac{1.5}{Z-0.75}+\frac{1.12Z-0.16}{Z^2-0.8Z+0.64}\right]=1.5\frac{Z}{Z-0.75}+1.12\frac{Z(Z-0.143)}{Z^2-0.8Z+0.64}$$

查表得.

$$y(kT)=1.5(0.75)^k+1.195(0.8)^k\cos(k\cdot 60°-20.4°)$$

据此即可计算出当 $k=0,1,2,\cdots$时，$y(kT)$的数值.

2° 长除法

如果 Z 变换象函数 $Y(Z)$以有理函数的形式给出，则可以直接通过用分子除以分母，得到无穷幂级数的展开形式. 如果所得级数是收敛的，则级数中的 Z^{-k}的系数就是时间序列中 $y(Tk)$的数值. 在用长除法求系数时，$Y(Z)$的分子与分母多项式都必须写成 Z^{-1}的升幂形式.

在上例中

$$Y(Z)=\frac{2.62-2.20Z^{-1}+1.08Z^{-2}}{1-1.55Z^{-1}+1.24Z^{-2}-0.48Z^{-3}}$$
$$=2.62+1.86Z^{-1}+0.72Z^{-2}+0.05Z^{-3}+\cdots$$

这个无穷级数是收敛的，逐项求反变换，得

$$y(0)=2.62,y(T)=1.86,y(2T)=0.72,y(3T)=0.05,\cdots$$

显然与查表所得数值一致. 不过一般得不到关于 $y(kT)$的通项表达式.

3° 反演积分的留数计算法(略).

五、差分方法

对于连续函数来说，微分、微商、微分方程是极其重要的基础知识．为了对离散信号和离散系统进行研究，必须引入差分、差商和差分方程．

设 y 是自变量 t 的函数，记为 $y(t)$．若取 Δt 作为 t 的一个步长，当 t 自某值 t_0 变化一个步长，即 $t=t_0+\Delta t$，y 的增量为 Δy，则称 Δy 是 y 在 $t=t_0$ 处的一阶差分，$\Delta y/\Delta t$ 称为差商．若记 $\Delta t=T$，记 $t=kT$ 处的函数值为 $y(kT)$，则 $t=kT$ 处的差分 $\Delta y(kT)=y[(k+1)T]-y(kT)$．

还可以引入一阶差分 $\Delta y(kT)$ 的差分，即 $\Delta[\Delta y(kT)]$，称为 y 在 $t=kT$ 处的二阶差分，记为 $\Delta^2 y(kT)$．

$$\begin{aligned}\Delta^2 y(kT)=\Delta[\Delta y(kT)]&=\Delta y[(k+1)T]-\Delta y(kT)\\&=y[(k+2)T]-y[(k+1)T]\\&\quad-[y(k+1)T]+y(kT)\\&=y[(k+2)T]-2y[(k+1)T]+y(kT)\end{aligned}$$

类似可写出 n 阶差分的表达式

$$\begin{aligned}\Delta^n y(kT)=&\,y(k+n)-C_n^1y[(k+n-1)T]+C_n^2y[(k+n-2)T]\\&+\cdots+(-1)^{n-1}C_n^{n-1}y[(k+1)T]+(-1)^n y(kT).\end{aligned}$$

不失一般性，可令 $T=1$，我们称形如

$$F(k,y(k),y(k+1),\cdots,y(k+n))=0$$

的方程为差分方程．线性差分方程的一般形式为

$$\begin{aligned}&y(k+n)+a_1y(k+n-1)+\cdots+a_ny(k+m)\\&\quad=b_1u(k+n-1)+\cdots+b_mu(k+m),\end{aligned}$$

式中，$y(k)$ 是待求的离散函数，$u(k)$ 是已知的离散函数，$a_i,b_i,i=1,2,\cdots,m$ 是已知系数．

与微分方程不同的是，差分方程的阶是用差分方程中 y 的最大附标 $(k+n)$ 和 y 的最小附标 $(k+m)$ 之差 $(n-m)$ 来规定的，例如

$$y(k+2)+y(k+1)+y(k-1)=u(k-1)$$

因为$(k+2)-(k-1)=3$，所以该差分方程是三阶．

求解线性差分方程主要有两种方法：递推法和 Z 变换法．

1．递推法

以一阶差分方程为例

$$y(k+1)-ay(k)=bu(k) \tag{2}$$

设 $y(0),u(0),u(1),\cdots$是已知的，将(1)式改写为

$$y(k+1)=a\lg(k)+bu(k)$$

设 $k=0$，则有 $y(1)=ay(0)+bu(0)$

设 $k=1$，则有 $y(2)=ay(1)+bu(1)=a^2y(0)+abu(0)+bu(1)$

依此递推，可写出解的一般表达式

$$y(k)=a^ky(0)+a^{k-1}bu(0)+a^{k-2}bu(1)+\cdots+bu(u-1)$$

对 u 阶线性方程，其稳定平衡的条件是特征方程 $\lambda^n+a_1\lambda^{n-1}+\cdots+a_n=0$ 的根 $\lambda_i(i=1,2,\cdots,n)$均有$|\lambda_i|<1$.

2．z 变换法(查表法)

用 z 变换法解差分方程的步骤与普通拉氏变换法解线性微分方程的步骤是相似的，即首先将原来的线性常系数差分方程进行 z 变换．得到以 z 变换式表示的代数方程，然后解出相应的变量，最后进行 z 反变换．

例 8-1　z 变换法解下列差分方程

$$y(k+2)+3y(k+1)+2y(k)=0, y(0)=0, y(1)=1$$

解　对上式取 z 变换，得

$$z^2y(z)-z^2y(0)-zy(1)+3zy(z)-3zy(0)+2y(z)=0$$

化简得，

$$y(z)=\frac{z}{z^2+3z+2}=\frac{z}{(z+1)(z+2)}=\frac{z}{z+1}-\frac{z}{z+2}$$

查表，$z[a^k]=\dfrac{z}{z-a}$，于是

$$z^{-1}[\frac{z}{z+1}]=(-1)^k, z^{-1}[\frac{z}{z+2}]=(-2)^k$$

故 $y(k)=(-1)^k-(-2)^k, k=0,1,2,\cdots$

§8.2 模型的参数估计

离散线性系统的数学模型是用差分方程表示的．我们面临的问题是，如何根据输入、输出的数据来确定差分方程．在已知模型的结构和阶数时，就是要估计差分方程的未知系数．因此是一个参数估计问题．估计的方法主要有最小二乘法，极大似然法，最大后验法和梯度法等等．无论从应用还是从理论的角度看，最小二乘法是最基本的．

设 n 阶单输入单输出线性离散系统的输入输出关系是

$$\begin{aligned}&y(k)+a_1y(k-1)+\cdots+a_ny(k-n)\\&=b_1u(k-1)+b_2u(k-2)+\cdots+b_nu(k-n)\end{aligned}\tag{1}$$

式中 $y(k),y(k-1),\cdots,y(k-n)$ 为输出信号在第 $k,k-1,\cdots,k-n$ 时刻的采样值；$u(k-1),u(k-2),\cdots,u(k-n)$ 是输入信号在第 $k-1,k-2,\cdots,k-n$ 时刻的采样值；$a_1,a_2,\cdots,a_n,b_1,b_2,\cdots,b_n$ 为系统的真实参数值．

为书写方便起见，引进位移算子：

$$y(k-l)=z^{-l}y(k),u(k-l)=z^{-l}u(k)\quad(l=1,2,\cdots,n)$$

于是(1)式可记为

$$y(k)[1+A(z^{-1})]=B(z^{-1})u(k)$$

其中，$A(z^{-1})=a_1z^{-1}+a_2z^{-2}+\cdots+a_nz^{-n}$

$$B(z^{-1})=b_1z^{-1}+b_2z^{-2}+\cdots+b_nz^{-n}$$

记 $\quad A=[a_1,a_2,\cdots,a_n]^T, B=[b_1,b_2,\cdots,b_n]^T.$

假设输入信号序列 $\{u(k)\}$ 是可以精确考察的，由于测量不精确或环境对过程的随机干扰使实际测量到的输出为

$$y(k)=-A(z^{-1})y(k)+B(z^{-1})u(k)+\xi(k)$$

其中，$\xi(k)$ 是一个由量测噪声所引起的随机变量，为简单起见，现在先假定它具有如下统计学特性：

1° 零均值 $E[\xi(k)]=0$

2° 对输入、输出信号是独立的

$$E[\xi(k)u(k-l)]=0, E[\xi(k)y(k-l)]=0 \quad (l=0,1,\cdots,k)$$

3° $\xi(k), k=0,1,2,\cdots$ 是一个不相关的随机变量序列

$$E[\xi(k)\xi(k-l)]=0 \quad (l=1,2,\cdots,k)$$

为了建立上述系统输入、输出关系的模型，不妨假设模型的结构已经确定为

$$y(k)+\hat{a}_1 y(k-1)+\cdots+\hat{a}_n y(k-n)$$
$$=\hat{b}_1 u(k-1)+\hat{b}_2 u(k-2)+\cdots+\hat{b}_n u(k-n)$$

其中，$\hat{a}_i, \hat{b}_i, i=1,2,\cdots,n$ 是模型的估计参数．由此系统输出的量测值和从上式求得的估计值的差 $e(k)$ 为

$$e(k)=y(k)+\hat{a}_1 y(k-1)+\hat{a}_2 y(k-2)+\cdots$$
$$+\hat{a}_n y(k-n)-\hat{b}_1 u(k-1)-\cdots-\hat{b}_n u(k-n)$$

称为模型的残差．采集 $n+N$ 组输入、输出数据 $y(k), u(k), k=1,2,\cdots,N+n$，构成以残差为输出的 N 个方程．

$$e(n+1)=y(n+1)+\hat{a}_1 y(n)+\hat{a}_2 y(n-1)+\cdots+\hat{a}_n y(1)$$
$$-\hat{b}_1 u(n)-\hat{b}_2 u(n-1)-\cdots-\hat{b}_n u(1)$$
$$e(n+2)=y(n+2)+\hat{a}_1 y(n+1)+\hat{a}_2 y(n)+\cdots+\hat{a}_n y(2)$$
$$-\hat{b}_1 u(n+1)-\hat{b}_2 u(n)-\cdots-\hat{b}_n u(2)-\cdots$$
$$e(n+N)=y(n+N)+\hat{a}_1 y(n+N-1)+\hat{a}_2 y(n+N-2)$$
$$+\cdots+\hat{a}_n y(N)-\hat{b}_1 u(n+N-1)-\cdots-\hat{b}_n u(N)$$

写成矩阵形式

$$e=Y-\varphi_N\hat{\theta}$$

其中 $\quad Y=[y(n+1), y(n+2), \cdots, y(n+N)]^T$

$$\varphi_N=\begin{bmatrix} y(1) & y(2) & \cdots & y(n) & u(1) & u(2) & \cdots & u(n) \\ y(2) & y(3) & \cdots & y(n+1) & u(2) & u(3) & \cdots & u(n+1) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ y(N) & y(N+1) & \cdots & y(N+n-1) & u(N) & u(N+1) & \cdots & u(N+n-1) \end{bmatrix}$$

$$\hat{\theta}=[-\hat{a}_n, -\hat{a}_{n-1}, \cdots, -\hat{a}_1, \hat{b}_n, \hat{b}_{n-1}, \cdots, \hat{b}_1]^T$$

$$e=[e(n+1), e(n+2), \cdots, e(n+N)]^T$$

我们采用残差序列的平方和作为衡量模型误差大小的指标，寻求模型参数 θ 使

$$\min J=\sum_{k=n+1}^{n+N} e^2(k)=\boldsymbol{e}^T\boldsymbol{e}$$

令 $\frac{\partial J}{\partial \theta}=\frac{\partial}{\partial \theta}\boldsymbol{e}^T\boldsymbol{e}=0$，即

$$\frac{\partial}{\partial \theta}(\boldsymbol{Y}-\boldsymbol{\varphi}_N\theta)^T(\boldsymbol{Y}-\boldsymbol{\varphi}_N\theta)=0$$

根据矢量函数对向量求导的乘法法则

$$\frac{\mathrm{d}(\boldsymbol{A}^T\boldsymbol{B})}{\mathrm{d}\boldsymbol{x}}=\frac{\mathrm{d}\boldsymbol{A}^T}{\mathrm{d}\boldsymbol{x}}\cdot\boldsymbol{B}+\frac{\mathrm{d}\boldsymbol{B}^T}{\mathrm{d}\boldsymbol{x}}\cdot\boldsymbol{A}$$

有

$$-2\boldsymbol{\varphi}_N^T(\boldsymbol{Y}+\boldsymbol{\varphi}_N\theta)=0$$

即 $\theta=(\boldsymbol{\varphi}_N^T\boldsymbol{\varphi}_N)^{-1}\boldsymbol{\varphi}_N^T\boldsymbol{Y}$.

最小二乘估计具有下述性质：

1° 若系统的随机噪声干扰序列 $\{\xi(k)\}$ 独立同分布，且 $E[\xi(k)]=0, E[\xi^2(k)]=\sigma^2$，则上述模型参数的最小二乘估计量是一致的，即

$$\hat{\theta}_N \xrightarrow{a.s} \theta(\text{当 } N\to\infty\text{时})$$

然而在通常情况下；噪声干扰都是相关的，这时最小二乘估计量 $\hat{\theta}_N$ 就不是一致估计．

2° 若系统的随机噪声干扰序列 $\{\xi(k)\}$ 对输出 $y(k)$ 是独立的，且均值为零，$E[\xi(k)]=0$，则模型参数的最小二乘估计是无偏的，即 $E[\hat{\theta}_N]=\theta$. 但在通常情况下，$\{\xi(k)\}$ 与 $\{y(k)\}$ 不相关这一假设是不成立的．因此，一般说来，最小二乘估计不是无偏的．不过，若 $\{\xi(k)\}$ 是不相关的，最小二乘就可以是强一致的，从而也是渐进无偏的．

3° 若系统的观测误差序列 $\{\xi(k)\}$ 同分布，均值 $E[\xi(k)]=0$，方差 $E[\xi^2(k)]=\sigma^2(k=1,2,\cdots,n)$，则估计误差的方差阵是 $\sigma^2(\phi_N^T\phi_N)^{-1}$；这可用来判断最小二乘估计参数的精度．

1. 最小二乘的递推算法

根据估计的一致性，样本容量越大时估计值接近真值的概率也越大．因此在实际工作中，为了提高参数估计值的精度总是希望能利用的数据尽量地多．但按最小二乘方法，每增加一个新数据，都得将全部的数据重新进行一次计算，这就要求把所有原来的数据保留起来．随着观测数据的增加必然造成计算机的存储量不能满足要求，而且计算量也逐渐变大，所以不能适应在线辨识的要求，因此现在广泛采用递推算法．

若用 $n+N$ 组数据求得 $\hat{\theta}_N$ 后，又取得一个新的观测数据 $y(n+N+1)$，$n(n+N+1)$，现在分析未知参数 $\hat{\theta}_{N+1}$的计算过程．

记

$$\varphi_{N+1}=\begin{bmatrix}\varphi_N\\ x_{n+N}^T\end{bmatrix}\qquad Y_{N+1}=\begin{bmatrix}Y_N\\ y(n+N+1)\end{bmatrix}$$

其中 $x_{n+N}^T=[y(N+1),y(N+2),\cdots,y(n+N),u(N+1),\cdots,u(n+N)]$，按照最小二乘法则，利用 $n+N+1$ 组观测数据的计算公式为

$$\hat{\theta}_{N+1}=[\phi_{N+1}^T\phi_{N+1}]^{-1}\phi_{N+1}^T Y_{N+1}$$

令
$$P_k=[\phi_k^T\phi_k]^{-1}\qquad(k=1,2,\cdots)$$

则
$$P_{N+1}=[\phi_{N+1}^T\phi_{N+1}]^{-1}$$

用 ϕ_{N+1}的表达式代入上式，得

$$P_{N+1}=\left\{[\varphi_N^T\quad X_{n+N}]\begin{bmatrix}\varphi_N\\ x_{n+N}^T\end{bmatrix}\right\}^{-1}=[\varphi_N^T\varphi_N+X_{n+N}X_{n+N}^T]^{-1}$$
$$=[P_N^{-1}+X_{n+N}X_{n+N}^T]^{-1}$$

利用矩阵求逆公式

$$(A+BC^T)^{-1}=A^{-1}-A^{-1}B(I+C^TA^{-1}B)^{-1}C^TA^{-1}$$

于是 $P_{N+1}=P_N-P_NX_{n+N}[1+X_{n+N}^TP_NX_{n+N}]^{-1}X_{n+N}^TP_N$，

则
$$\hat{\theta}_{n+1}=P_{N+1}[\varphi_N^T\cdots X_{n+N}]\begin{bmatrix}Y_N\\ y(n+N+1)\end{bmatrix}$$
$$=P_{N+1}[\varphi_N^TY_N+X_{n+N}y(n+N+1)]$$
$$=[P_N-P_NX_{n+N}(1+X_{n+N}^TP_NX_{n+N})^{-1}X_{n+N}^TP_N]$$

$$
\cdot\left[\varphi_N^T Y_N + X_{n+N}\, y(n+N+1)\right]
$$

$$
= P_N \varphi_N^T Y_N - P_N X_{n+N}(1+X_{n+N}^T P_N X_{n+N})^{-1} X_{n+N}^T P_N \varphi_N^T Y_N
$$

$$
+ P_N X_{n+N}\, y(n+N+1) - P_N X_{n+N}(1 + X_{n+N}^T P_N X_{n+N})^{-1} X_{n+N}^T P_N X_{n+N}\, y(n+N+1)
$$

注意到 $P_N \varphi_N^T Y_N = (\varphi_N^T \varphi_N)^{-1} \varphi_N^T Y_N = \hat{\theta}_N$

代入得

$$
\begin{aligned}
\hat{\theta}_{N+1} &= \hat{\theta}_N - P_N X_{n+N}(1+X_{n+N}^T P_N X_{n+N})^{-1} X_{n+N}^T \hat{\theta}_N \\
&\quad + P_N X_{n+N}(1+X_{n+N}^T P_N X_{n+N})^{-1}(1+X_{n+N}^T P_N X_{n+N}) \\
&\quad \cdot y(n+N+1) - P_N X_{n+N}(1+X_{n+N}^T P_N X_{n+N})^{-1} X_{n+N}^T P_N X_{n+N} \\
&\quad \cdot y(n+N+1) \\
&= \hat{\theta}_N - P_N X_{n+N}(1+X_{n+N}^T P_N X_{n+N})^{-1} X_{n+N}^T \hat{\theta}_N \\
&\quad + P_N X_{n+N}(1+X_{n+N}^T P_N X_{n+N})^{-1}(1+X_{n+N}^T P_N X_{n+N} \\
&\quad - X_{n+N}^T P_N X_{n+N})\, y(n+N+1) \\
&= \hat{\theta}_N - P_N X_{n+N}(1+X_{n+N}^T P_N X_{n+N})^{-1} X_{n+N}^T \hat{\theta}_N \\
&\quad + P_N X_{n+N}(1+X_{n+N}^T P_N X_{n+N})\, y(n+N+1) \\
&= \hat{\theta}_N - P_N X_{n+N}(1+X_{n+N}^T P_N X_{n+N})^{-1}\left[y(n+N+1) - X_{n+N}^T \hat{\theta}_N\right]
\end{aligned}
$$

利用上述递推公式，当取得一个新的观测数据 $y(n+N+1)$ 时，用历史上的 n 次数据 $y(N+1), y(N+2), \cdots, y(N+n)$，以及上一次的系数估计值 $\hat{\theta}_N$ 即可得到新的系数估计值 $\hat{\theta}_{N+1}$.

2. 最小二乘估计的缺陷及其改进

在推导基本最小二乘公式时，为简单起见，把残差 $e(k)$ 看作是在 k 时刻上的量测噪声所引起的，是不相关序列，因此得到最小二乘估计量的无偏性．然而这个假设是不符合实际情况的．因为模型在 k 时刻的估计值是由 k 时刻以前各时刻的输出、输入的量测值 $y(k-1), y(k-2), \cdots, y(k-n), u(k-1), \cdots, u(k-n)$ 计算出来的，而在上述每个时刻上的量测值和 k 时刻上的量测值 $y(k)$ 一样，也都含有相应于该时刻的量测噪声．于是，在 k 时刻上的残差 $e(k)$ 也必然与 k 时刻以前各时刻的残差 $e(k-1), e(k-2), \cdots$，相关．所以最小二乘的参数估计值 $\hat{\theta}$ 并不依概率收敛于真值 θ. 为

解决这一问题，人们提出了不少改进的方法.

广义最小二乘法，其基本思路是用噪声滤波的方法把一个具有相关残差序列$\{e(k)\}$的模型化成一个等效的具有不相关残差序列$\{\xi(k)\}$的模型，再用最小二乘方法实现参数的无偏估计.

其他方法还有辅助变量法、极大似然估计法以及相关分析和最小二乘的两步法等等.

例 8-2 局部脑血流量测定.

用放射性同位素测量大脑局部血流量的方法如下：由受试者吸入含有某种放射性同位素的气体，然后将探测器置于受试者头部某固定处，定时测量该处的放射性记数率(简称记数率)，同时测量他呼出气的记数率.

由于动脉血将肺部的放射性同位素输送至大脑，使脑部同位素增加，而脑血又将同位素带离，使同位素减少. 实验证明由脑血流引起局部区域计数率下降的速率与当时该区域的记数率成正比，其比例系数反映该处的脑血流量，被称为脑血流量系数，只要确定该系数即可推算出脑血流量. 动脉血从肺输送同位素至大脑引起脑部记数率上升的速率与当时呼出气的记数率成正比.

若其受试者的测试数据如表 8-1.

试建立确定脑血流系数的数学模型并计算上述受试者的脑血流系数(材料取自上海 1990 年数学模型竞赛试卷). 假定：

1° 脑部记数率上升只与肺部的放射性同位素有关，上升速率与呼出气的记数率成正比；

2° 脑部记数率下降只与该处脑血流量有关，其下降速率正比于脑记数率，这里忽略了放射性元素的衰变和其他因素；

3° 脑血流量在测量期间恒定. 心脏博动、被试者大脑活动、情感波动等带来的变化可忽略；

4° 每次仪器测量为相互独立事件，各测量值无记忆关联；

5° 放射性同位素在人体内传递从吸入气体(含有放射物)开始，假定一次吸入，则认为瞬时在肺中达到最大浓度；

表 8-1　局部脑血流量测试数据

时间	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75		
头部记数率	1534	1528	1468	1378	1272	1162	1052	947		
呼出气记数率	2231	1534	1054	724	498	342	235	162		
时间	3.00	3.25	3.50	3.75	4.00	4.25	4.50	4.75	5.00	5.25
头部记数率	348	757	674	599	531	471	417	369	326	288
呼出气记数率	111	76	52	36	25	17	12	8	6	4
时间	5.50	5.75	6.00	6.25	6.50	6.75	7.00	7.25	7.50	7.75
头部记数率	255	225	199	175	155	137	121	107	94	83
呼出气记数率	3	2	1	1	1	1	0	0	0	0
时间		8.00	8.25	8.50	8.75	9.00	9.25	9.50	9.75	10.00
头部记数率		73	65	57	50	44	39	35	31	27
呼出气记数率		0	0	0	0	0	0	0	0	0

6° 吸入气体脑时，脑中放射性记数率为零；

7° 脑血流量与脑血流系数成单值函数关系，求得后者即可确定前者．

记头部记数率为 $h(t)$．设某时刻 $t_0\geqslant 0$ 时，头部记数率为 h_0，Δt 时刻以后，记数率变为 $h_0+\Delta h$，由题设和假定，Δh 仅与三个因素有关：

1° 肺动脉血将肺部的放射性同位素送到大脑，使脑部记数率增加 Δh_1；

2° 脑血流将同位素带离，脑记数率下降 Δh_2；

3° 放射性元素自身有衰减，设其半衰期为 τ，由此引起的记数率下降为 Δh_3．

又由医学实验和假定有

$$\frac{\mathrm{d}h_1}{\mathrm{d}t}=\beta p(t),$$

其中，$p(t)$为呼出气体记数率；

$$\frac{\mathrm{d}h_2}{\mathrm{d}t}=\alpha h(t)$$

以及

$$\Delta h_3=-h(t)(\frac{1}{2})^{\frac{\Delta t}{\tau}}\cdot \ln 2\cdot \frac{1}{\tau}\Delta t+0(\Delta t)$$

$$\frac{\mathrm{d}h_3}{\mathrm{d}t}=-\frac{\ln 2}{\tau}h(t)$$

所以有

$$\frac{\mathrm{d}h}{\mathrm{d}t}=-\alpha h+\beta p-\frac{\ln 2}{\tau}h$$

其中，α是脑血流系数，β是呼出气体记数率系数.

由于在测试时放射性同位素的半衰期一般很大，这样会给测量和试验带来严重影响，因此假定$\tau\to\infty$，于是有

$$\frac{\mathrm{d}h}{\mathrm{d}t}=-\alpha h(t)+\beta p(t) \tag{2}$$

首先对$p(t)$作最小二乘拟合，得

$$\ln p(t)=9.1628-1.2807t$$

其相关系数$r=0.9999$，因此可以认为$p(t)$是负指数曲线$p(t)=Ae^{-\lambda t}$，其中$A=e^{9.1648}$，$\lambda=1.4807$. 由(1)式及假定(6)$h(0)=0$，得

$$h(t)=\frac{\beta A}{\alpha-\lambda}(e^{-\lambda t}-e^{-\alpha t})$$

(可用非线性最小二乘对参数进行估计)

一、算法模型 I

将(1)式离散化，记时间间隔为T，利用向前差商公式，得

$$\frac{h_{n+1}-h_n}{T}=-\alpha h_n+\beta p_n$$

或

$$h_{n+1}=(1-\alpha T)h_n+\beta T p_n \tag{3}$$

以及向后差商公式，得

$$\frac{h_n-h_{n-1}}{T}=-\alpha h_n+\beta p_n$$

用差分方法求解，截断误差为 $O(T^2)$，为提高精度，可用三次样条插值在每两个结点的中点进行插值，缩短步长，使截断误差减小到原来的 $\frac{1}{4}$.

二、算法模型 I

对(1)式作拉普拉斯变换，得

$$sh(s)-h(0)=-\alpha h(s)+\beta p(s)$$

其中，$h(0)=0$，求该系统的传递函数

$$G(s)=\frac{h(s)}{p(s)}=\frac{\beta}{s+\alpha}$$

对 $p(s)$进行采样，采样周期为 T，由于采样得到的是脉冲量，加入一个保持器.

若取零阶保持器 $H_1(s)=\frac{1-e^{-Ts}}{s}$，则整个系统的广义传递函数

$$G_1(s)=\frac{1-e^{-Ts}}{s}\cdot\frac{\beta}{s+\alpha}$$

对 $G_1(s)$作 z 变换，

$$z(G_1(s))=G_1(z)=\frac{h(z)}{p(z)}=\frac{\beta}{\alpha}\,\frac{1-e^{-T\alpha}}{z-e^{-T\alpha}}$$

或

$$zh(z)-e^{-T\alpha}h(z)=\frac{\beta}{\alpha}(1-e^{-T\alpha})p(z)$$

得到采用零价保持器离散化差分方程为

$$h_{n+1}=e^{-\alpha T}h_n+\frac{\beta}{\alpha}(1-e^{-T\alpha})p_n \tag{4}$$

将 $e^{-\alpha T}$展开，取前两项，有

$$e^{-\alpha T}=1-\alpha T+0(T)$$

略去高次项，即得

$$h_{n+1}=(1-\alpha T)h_n+\beta T p_n$$

这正好是用一阶差分法离散微分方程后的结果．由此可见，用零阶保持器 z-变换方法得到了一个含有 T 的所有高次项的高精度模型．

如果采用三角形保持器 $H_2(s)=\frac{e^{Ts}(1-e^{-Ts})^2}{Ts^2}$，因三角形保持器对状态的预测不仅用了当前时刻的采样值，而且还以线性外推了两采样点间的值，其差分程为

$$h_{n+1}=e^{-\alpha T}h_n+\frac{\beta}{T\alpha^2}(e^{-\alpha T}-1+T\alpha)p_{n+1}$$
$$-\frac{\beta}{T\alpha^2}(T\alpha e^{-\alpha T}+e^{-\alpha T}-1)p_n$$

三、用递推最小二乘法实现模型参数的估计

将(3)式多个时刻的方程写成矩阵形式

$$\begin{pmatrix}h_k\\h_{k+1}\\h_{k+2}\end{pmatrix}=\begin{pmatrix}h_{k-1}&p_{k-1}\\h_k&p_k\\h_{k+1}&p_{k+1}\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$$

其中，$x=e^{-\alpha T}$，$y=\frac{\beta}{\alpha}(1-e^{-\alpha T})$

记作 $H=\varphi\,\theta$，采用最小二乘法

$$\hat{\theta}=[\varphi^T\varphi]^{-1}\varphi^T H$$

在上面过程中，若 h_k，p_k 这些数据点过多，φ 矩阵的规模也随之增大，而另一方面也必须尽可能多地利用所有数据，故改进递推算法：

$$\begin{cases}q_{n+1}=q_n-q_n\varphi_{n+1}[1+\varphi_{n+1}^T q_n\varphi_{n+1}]^{-1}\varphi_{n+1}^T q_n\\ k_n=q_n\varphi_{n+1}/(1+\varphi_{n+1}^T q_n\varphi_{n+1})\\ \theta_{n+1}=\theta_n+R_n(h_{n+1}-\varphi_{n+1}^T\theta_n)\end{cases}$$

其中，$\varphi_{n+1}=[h_{n+1},p_{n+1}]^T$ 为新的观测，$\theta_n=[\alpha_n,\beta_n]^T$ 为前一时刻的拟合值，初值为

$$\theta_0=[0,0]^T,q_0=\begin{bmatrix}10^5&0\\0&10^5\end{bmatrix}$$

四、测试和结论

各种算法的测试结果以及求得的结果见表 8-2.

表 8-2 四种算法比较

算法 \ 结果	算法测试		问题结果 (α)
	设定值	拟合值	
1. 直接前插分	1.0000	0.8848	0.4700
2. 样条插值、再作 1	——	——	0.4835
3. z 变换法(零阶保持器)	1.0000	1.0000	0.5000
4. z 变换法(三角形保持器)	1.0000	1.0000	0.5001
5.			

以上测试数据均在第二步后收敛到稳定值，问题的结果数据经过若干步波动后达到终值，说明算法具有稳定性和快速收敛性. 从表 8-1 中的数据可以看到截断误差对结果的影响及用 z 变换离散化方法的优越性和准确性，并且得到 α 的最佳拟合值为 0.500.

该问题亦可利用 § 7.3 方法对数据直接进行处理，留给读者作为练习.

§ 8.3 线性模型结构的辨识

从线性离散模型的一般形式

$$y(k)+a_1y(k-1)+\cdots+a_ny(k-n)$$
$$=b_1u(k-d-1)+\cdots+b_nu(k-d-n)$$

可以看出，线性模型的结构参数是指模型的阶数 n 和纯延迟时间 d.

在模型的参数估计中假定模型的结构参数为已知，但一个实际过程的模型结构事先往往是不知道的，所以在参数估计前应先

辨识模型结构．

一、纯延迟时间 d 的确定

这里讨论用参数估计的方法来确定纯延迟时间 d.

设 n_0 为系统的真阶，d_0 为系统的真纯延迟时间．为确定纯延迟 d_0，可先任意假设一个相当大的阶 n（例如 $n \geqslant n_0+d_0$）构造模型

$$
\begin{aligned}
& y(k)+\hat{a}_1 y(k-1)+\cdots+\hat{a}_n y(k-n) \\
& \quad =\hat{b}_1 u(k-1)+\cdots+\hat{b}_r u(k-r)+\hat{b}_{r+1} u(k-r-1) \\
& \qquad +\cdots+\hat{b}_n u(k-n)
\end{aligned}
$$

式中 $\hat{a}_i, \hat{b}_i \quad i=1,2,\cdots,n$ 为估计参数．

若得到的估计量 $\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_r$ 等数值很小，几乎可以忽略，而且有 $|\hat{b}_{r+1}| \gg |\hat{b}_r|$，则可认为 r 就是纯延迟时间 d_0. 这一点可作如下解释：由于过程具有纯延迟 d_0，在 $k-d_0-1$ 以后瞬时的输入都不会引起 $y(k)$ 的响应，现在由试验数据得到的估计量 $\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_r$ 的数值几乎可忽略，而且 $|\hat{b}_{r+1}| \gg |\hat{b}_r|$，这意味着测试数据 $y(k)$ 不是由 $u(k-1), u(k-2), \cdots, u(k-r)$ 所引起的响应，而是由 $u(k-r-1), u(k-r-2), \cdots, u(k-n)$ 引起的响应，于是 r 就可认为是系统的纯延迟时间 d_0.

二、模型阶数的检验

模型阶数的检验有很多种方法，常用的有如下几类：零极点对消检验、损失函数检验、残差检验、偏相关函数法、AIC 准则以及 F 检验法等等．其中 F 检验法的基本思想是：若在所选的阶数 n_1 和 n_2 下进行参数估计所得到的残差平方和记为 $J(n_1)$ 和 $J(n_2)$，如果残差是独立正态分布的随机变量，则当 $n_2>n_1 \geqslant n_0$ 时（n_0 为理想的阶），$J(n_2)$ 和 $[J(n_1)-J(n_2)]$ 也是独立的随机变量，定义 $t=\dfrac{J(n_1)-J(n_2)}{J(n_2)}$、$\dfrac{N-2n_2}{2(n_2-n_1)}$，当采样点数 N 很大时，随机变量 t 是渐近 $F[2(n_2-n_1),(N-2n_2)]$ 分布．若给定一个置信度 α，通常取 α

$=0.01$ 或 0.05，从 F 分布函数表中可以找到 t_a，若 $t<t_a$，则表示 $J(n_2)$的概率小于 $J(n_1)$，则 n_1 就是被估计系统的阶；若 $t\geqslant t_a$，则表示 $n_2>n_1\geqslant n_0$ 不成立，需要增加阶次，继续运算，直至 $t<t_a$ 为止．

§8.4 模型的选择

时间序列的线性模型在系统辨识中占有重要地位，线性模型有多种形式，模型辨识的任务就是找出某一时间序列最合适的数学模型并决定其阶数．由于实际问题中大多数的平稳时间序列其期望值常常不为零，为了数学上处理方便起见，可作如下变换：

$$w_t=x_t-\mu \qquad (t=1,2,\cdots)$$

其中，$u=E(x_t)$，则$\{w_t\}(t=1,2,\cdots)$为零均值平稳时间序列．

通常在实际工程中，平稳时间序列 w_t 的数学模型可以表示为下面三种形式．

一、自回归模型(AR)模型

任何一个时刻 t 的数值 w_t 可表示为 p 个时刻的数值 w_{t-1}，w_{t-2}，…，w_{t-p}的线性组合，再加上 t 时刻的白噪声 n_t，即

$$w_t=a_1w_{t-1}+a_2w_{t-2}+\cdots+a_pw_{t-p}+n_t$$

即 w_t 由该时刻以前的各个值的各权和表示，由于拟合误差 n_t 是随机的，互相独立的，互不相关的，所以 n_t 具有白噪声的特性，p 为模型阶数。

二、滑动平均模型(MA 模型)

对于一个平稳时间序列$\{w_t\}$，存在一个白噪声序列 $n_1,n_2,\cdots,n_t,\cdots$，可以使 t 时刻的数值 w_t 表示成此时刻白噪声，减去前面 q 个时刻的白噪声 $n_{t-1},n_{t-2},\cdots,n_{t-q}$的加权平均

$$w_t=n_t-b_1n_{t-1}-\cdots-b_nn_{t-q}$$

其中，q 为阶数．

三、自回归滑动平均模型(ARMA 模型)

ARMA 模型也称为混合模型,其数学表达式为

$$w_t = a_1 w_{t-1} + a_2 w_{t-2} + \cdots + a_p w_{t-p} + n_t - b_1 n_{t-1} \cdots - b_q n_{t-q}$$

式中 p 及 q 为模型的阶数.

为了讨论模型的辨识方法,需要引入自相关函数和偏相关函数.

自协方差函数 r_k 定义为

$$r_k = E[(x_t - \mu)(x_{t+k} - \mu)] = E(w_t w_{t+k}),$$

其中,k 表示时间延迟量. 自相关函数定义为

$$\rho_k = E\left[\frac{x_t - \mu}{\sigma} \cdot \frac{x_{t+k} - \mu}{\sigma}\right] = \frac{r_k}{\sigma^2} = \frac{r_k}{r_0}$$

样本自协方差函数 $\hat{r}_k$ 可取为

$$\hat{r}_k = \sum_{j=1}^{n-k} w_j w_{j+k} / (n-k) \qquad (k = 0, 1, \cdots, k)$$

当 n 相当大,k 相对很小时,上式可写成

$$\hat{r}_k = \sum_{j=1}^{n-k} w_j w_{j+k} / n$$

样本自相关函数则为

$$\hat{\rho}_k = \frac{\hat{r}_k}{\hat{r}_0} \qquad (k = 0, 1, \cdots, K)$$

在实际计算中,一般取 $n > 50$,ρ_k 的个数 $K < \frac{n}{4}$. 通常 K 值为 $\frac{n}{10}$ 左右.

为研究自相关函数 ρ_k 的特性,考虑 AR 模型

$$w_t = a_1 w_{t-1} + a_2 w_{t-2} + \cdots + a_p w_{t-p} + n_t$$

两边同时乘以 w_{t-k},并求期望. 显然对 $k > 0$,等式右边最后一项有 $E(w_{t-k} - n_t) = 0$(实因 n_t 只对 w_t 有影响,而与以前的值 w_{t-k} 无关),得

$$r_k = a_1 r_{k-1} + a_2 r_{k-2} + \cdots + a_p r_{k-p}.$$

将上式两边除以 r_0，则

$$\rho_k=a_1\rho_{k-1}+a_2\rho_{k-2}+\cdots+a_p\rho_{k-p}\quad(k>0)$$

这是一个差分方程，按照 z 变换法求解差分方程的理论，该差分方程的解具有下列形式

$$\rho_k=A_1G_1^k+A_2G_2^k+\cdots+A_pG_p^k$$

若 G_i 全为实数，则 ρ_k 指数衰减，其快慢取决于主根；若 G_i 中有复根，则 ρ_k 振荡衰减．

自相关函数 ρ_k 在 $k>q$ 时全为零的性质称为 q 步截尾性．自相关函数不能在某步之后截尾，而是随着 k 增大逐渐衰减，但受负指数函数控制，这种特性称为拖尾性．

上述结果对 $AR(1)$ 模型只涉及"一步相关性"，为什么 ρ_k 在 $k=1$ 之后不截尾，而当 $k\to\infty$ 才逐渐趋于零，这是因为 w_t 与 w_{t-1} 有关，而 w_{t-1} 又与 w_{t-2} 有关，…，如此类推，w_t 和 w_{t-3}，w_{t-4}，…也都有关系．这种情况说明了用简单的相关函数来说明两个变量之间内在的相关性，有时会导致错误的印象．因为假如两个变量之间有正的自相关，其原因可能有两种：一种是一个变量的增加导致另一个变量的增加，另一种是某一个（或多个）变量的变化使被考察的两个变量都增加了．因此还需要有另一种表达两个变量之间相关性的方法，它能够排除其他变量的影响，这就是偏相关函数．可以设想，如果在排除了"局外"变量的影响之后，两个变量本身实际上并不存在相关性，就说明看起来似乎密切的相互关系是由于其他变量作用的缘故．反之，假若两个变量的相关值小于偏相关值，那么必定是其他变量"冲淡"了这两个变量间的相互关系．

根据以上分析，如果有三个随机变量 u,v,w，其联合概率密度为 $p(u,v,w)$，在 w 给定条件下，u 和 v 的条件概率密度记作 $p(u,v|w)=p(u,v,w)/p(w)$．则 w 给定时，u 和 v 的偏相关定义为

$$\rho_{u,v|w}=\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(u-\bar{u})(v-\bar{v})p(u,v|w)\mathrm{d}u\mathrm{d}v}{(\mathrm{Var}[u]\mathrm{Var}[v])^{1/2}}$$

或

$$\rho_{u,v|w}=\frac{E[(u-\bar{u})(v-\bar{v})]}{(\mathrm{var}[u]\mathrm{Var}[v])^{1/2}}$$

式中 E 表示关于条件密度函数 $p(u,v|w)$ 的条件期望．

在时间序列模型中，为了方便，定义 k 步迟后的偏相关为 $k-1$ 个中间值 $w_{t-1},w_{t-2},\cdots,w_{t-(k-1)}$ 一定时，w_t 和 w_{t-k} 之间的条件相关，记为

$$\rho_{w_t,w_{t-k}|w_{t-1},w_{t-2},\cdots,w_{t-(k-1)}}=\frac{E[w_tw_{t-k}]}{\sqrt{E[w_t^2]E[w_{t-k}^2]}}$$

$$=\frac{E[w_tw_{t-k}]}{\mathrm{var}[w_t]}.$$

现在设 $\{w_t\}$ 为零均值平稳序列，由 $w_{t-1},\cdots,w_{t-k}$，对 w_t 作最小方差估计．

$$w_t=\alpha_{k_1}w_{t-1}+\alpha_{k_2}w_{t-2}+\cdots+\alpha_{kk}w_{t-k}+n_t$$

两边乘以 $w_{t-k}(k>0)$，并取条件期望

$E[w_tw_{t-k}]=\alpha_{k1}E[w_{t-1}w_{t-k}]+\cdots+\alpha_{kk}E[w_{t-k}^2]+E[n_tw_{t-k}]$ 因为 $w_{t-1},w_{t-2},\cdots,w_{t-(k-1)}$ 是既定条件，可提到条件期望符号之外，则

$$E[w_tw_{t-k}]=\alpha_{k1}w_{t-1}E[w_{t-k}]-\cdots+\alpha_{kk}E[w_{t-k}^2]+E[n_tw_{t-k}]$$

注意到 $E[w_t]=0$，且 $k>0$ 时，n_t 和 w_{t-k} 不相关．

$$E[w_tw_{t-k}]=\alpha_{kk}\mathrm{Var}[w_t]$$

所以

$$\alpha_{kk}=\frac{E[w_tw_{t-k}]}{\mathrm{Var}[w_t]}.$$

可见偏相关函数值 $\alpha_{kk}(k=0,1,2,\cdots)$ 就是按 k 阶自回归方程对 w_t 作线性最小方差估计时的第 k 项（最后一项）系数．显然 $\alpha_{00}=1,\alpha_{11}=\rho_1$．由于 $\alpha_{kj},j=1,2,\cdots,k$，是基于最小方差原则，由

$$J=E[w_t-\sum_{j=1}^{k}\alpha_{kj}w_{t-j}]^2$$

令 $\partial J/\partial\alpha_{kj}=0,j=1,2,\cdots,k$，可得

$$\begin{bmatrix}\hat{\alpha}_{k1}\\ \hat{\alpha}_{k2}\\ \cdots\\ \hat{\alpha}_{kk}\end{bmatrix}=\begin{bmatrix}1 & \rho_1 & \cdots & \rho_{k-1}\\ \rho_1 & 1 & \cdots & \rho_{k-2}\\ \cdots & \cdots & \cdots & \cdots\\ \rho_{k-1} & \rho_{k-2} & \cdots & 1\end{bmatrix}^{-1}\begin{bmatrix}\rho_1\\ \rho_2\\ \cdots\\ \rho_k\end{bmatrix}$$

该方程称为尤尔-沃克(Yule-Walker)方程．凡满足上述方程的 $\alpha_{11},\alpha_{22},\cdots,\alpha_{kk}$ 为 w_t 的偏相关函数．

用上面公式求偏相关函数计算量较大，用下面的递推公式则要方便得多．在实用中

$$\hat{\alpha}_{11}=\rho_1 \tag{1}$$

$$\hat{\alpha}_{k+1,k+1}=(\rho_{k+1}-\sum_{j=1}^{k}\rho_{k+1-j}\hat{\alpha}_{kj})(1-\sum_{j=1}^{k}\rho_j\alpha_{kj})^{-1} \tag{2}$$

$$\hat{\alpha}_{k+1,j}=\hat{\alpha}_{kj}-\hat{\alpha}_{k+1,k+1}\hat{\alpha}_{k,k-(j-1)}\quad(j=1,2,\cdots,k) \tag{3}$$

递推的顺序是

$$\hat{\alpha}_{11}\xrightarrow[k=1]{\text{用}(B)}\hat{\alpha}_{22}\xrightarrow[k=1,j=1]{\text{用}(C)}\hat{\alpha}_{21}\xrightarrow[k=2]{\text{用}(B)}\hat{\alpha}_{33}\xrightarrow[k=2,j=1]{\text{用}(C)}\hat{\alpha}_{31}\xrightarrow[k=2,j=2]{\text{用}(C)}\hat{\alpha}_{32}\xrightarrow[k=3]{\text{用}(B)}\hat{\alpha}_{44}\rightarrow\cdots$$

为研究 AR 模型的偏相关函数的特性，考虑

$$\begin{aligned}J&=E(w_t-\sum_{j=1}^{k}a_{kj}w_{t-j})^2\\&=E(a_1w_{t-1}+a_2w_{t-2}+\cdots+a_pw_{t-p}+n_t-\sum_{j=1}^{k}a_{kj}w_{t-j})^2\\&=E[n_t+\sum_{j=1}^{p}(a_j-a_{kj})w_{t-j}-\sum_{j=p+1}^{k}a_{kj}w_{t-j}]^2\\&=\sigma^2+E[\sum_{j=1}^{p}(a_j-a_{kj})w_{t-j}-\sum_{j=p+1}^{k}a_{kj}w_{t-j}]^2\end{aligned}$$

当 $k>p$ 时，用如下取值方法可使 J 达到最小

$$\alpha_{kj}=\begin{cases}\alpha_j, & 1\leqslant j\leqslant p\\ 0, & p+1\leqslant j\leqslant k,k=p,p+1,p+2,\cdots\end{cases}$$

这时括号内第一项由于 $1\leqslant j\leqslant p$ 时 $\alpha_{kj}=\alpha_j$，而第二项由于 $p+1\leqslant j\leqslant k$ 时 $\alpha_{kj}=0$，特别在 $j=k$ 而 $k=p+1,p+2,\cdots$ 时，$\alpha_{kk}=0$. 这说

明 AR 模型在 $k\approx p$ 处开始截尾．

用类似方法对 MA(q)模型、ARMA(p,q)模型进行研究，结果如下表 8-3.

表 8-3　三种模型的 ρ_k 与 α_{kk} 特性

模型		AR(p)	MA(q)	ARMA(p,q)
相关数	ρ_k	拖尾	截尾	拖尾
	α_{kk}	截尾	拖尾	拖尾

在实际处理中，因为 $\hat{\rho}_k\approx\rho_k$，$\hat{\alpha}_{kk}\approx\alpha_{kk}$，当$|\hat{\rho}_k|$或$|\alpha_{kk}|$很小时可以认为 $\rho_k=0$ 或 $\alpha_{kk}=0$. 一般若在 $k>q$ 时，$|\rho_k|<\frac{2}{\sqrt{n}}$或$|\hat{\alpha}_{kk}|<\frac{2}{\sqrt{n}}$，可以认为 $\hat{\rho}_k$ 或 $\hat{\alpha}_{kk}$截尾，截在 $k=q$ 处．

例 8-3　某化学反应过程温度记录 200 个数据，计算得样本自相关系数和样本偏相关函数如表 8-4 所示．

表 8-4　样本自相关和偏相关函数表

k	1	2	3	4	5	6	7	8
$\hat{\rho}_k$	−0.73	−0.84	−0.13	−0.11	−0.01	−0.04	0.09	−0.05
$\hat{\alpha}_{kk}$	−0.73	−0.64	−0.71	−0.82	−0.73	−0.75	−0.76	−0.72
k	9	10	11	12	13	14	15	
$\hat{\rho}_k$	−0.08	0.13	−0.04	0.07	−0.05	0.02	0.03	
$\hat{\alpha}_{kk}$	−0.14	−0.32	0.11	−0.16	−0.12	−0.10	−0.07	

分别作点图，如图 8-3 所示．

由图可见，$\hat{\alpha}_{kk}$拖尾．此时$\frac{2}{\sqrt{n}}=\frac{2}{\sqrt{200}}\approx0.16$，当 $k>3$ 时，$|\rho_k|<0.16$，所以 $\hat{\rho}_k$ 截尾，截在 $k=3$ 处．因此可以认为线性模型是三阶滑动平均模型 MA (3)．

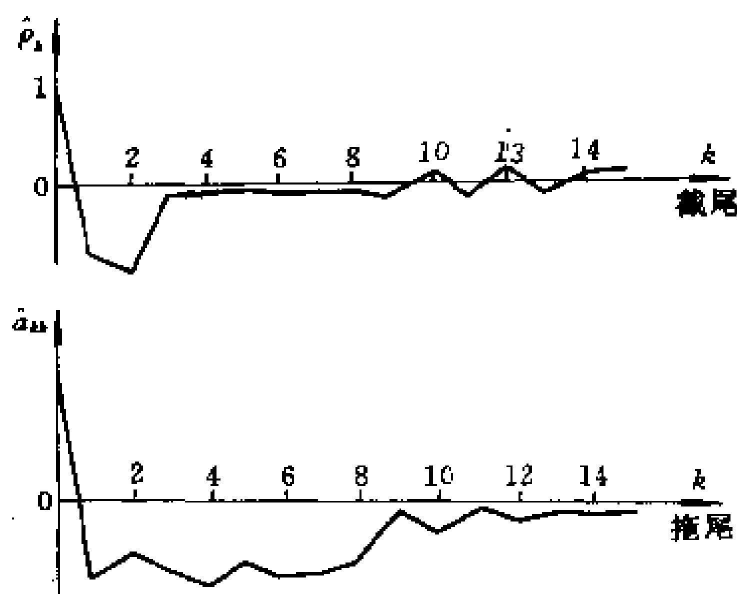


图 8-3 截尾与拖尾

例 8-4 根据某人心跳时间间隔所得 400 个数据，算得其样本自相关函数与样本偏相关函数见表 8-5，分别画点图（如图 8-4）.

表 8-5 样本自相关与偏相关函数表

k	1	2	3	4	5	6	7	8
$\hat{\rho}_k$	0.57	0.47	0.44	0.47	0.45	0.38	0.33	0.37
$\hat{a}_{kk}$	0.57	0.22	0.16	0.20	0.11	0.01	−0.03	0.10
k	9	10	11	12	13	14	15	
$\hat{\rho}_k$	0.39	0.42	0.32	0.31	0.27	0.25	0.24	
$\hat{a}_{kk}$	0.09	0.13	−0.06	−0.02	−0.06	−0.07	0.01	

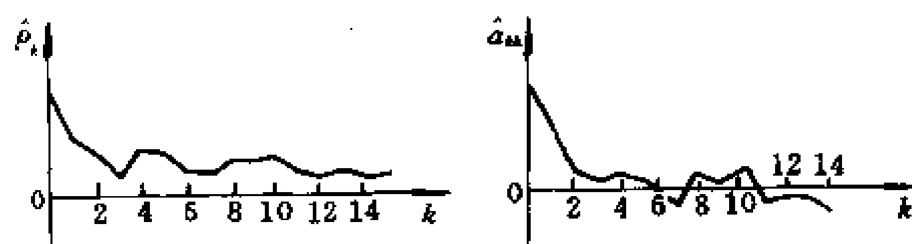


图 8-4 心跳 $\hat{\rho}_k$ 和 $\hat{A}_{kk}$ 点图

因为 $\frac{2}{\sqrt{n}}=\frac{2}{\sqrt{400}}=0.1$，数值较小，所以 $\hat{\rho}_k$ 与 $\hat{\alpha}_{kk}$ 都看成拖尾. 线性模型可以看成 ARMA 模型，不过模型的阶数还需用其他方法确定.

有时会遇到样本自相关函数与样本偏相关函数至少有一个既不是拖尾又不是截尾.

例 8-5 某化学反应过程温度记录数据，算出 $\hat{\rho}_k$ 和 $\hat{\alpha}_{kk}$ 分别作图，如图 8-5 所示.

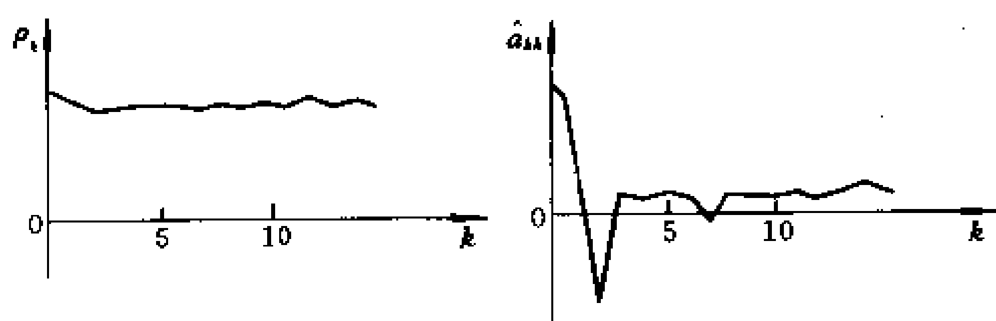


图 8-5　化学反应过程中温度 $\hat{\rho}_k$ 和 $\hat{\alpha}_{kk}$ 点图

由图可见，$\hat{\alpha}_{kk}$ 是截尾的，而 $\hat{\rho}_k$ 既不是截尾又不是拖尾的，此时线性模型不是前面指出的三种. 可以认为这个化学反应过程的温度是非线性的，或是非平稳时间序列.

用上述是否截尾或拖尾等来判别模型及其阶数只是近似的，有时画出的 $\hat{\rho}_k$ 及 $\hat{\alpha}_{kk}$ 很难看出是截尾还是拖尾，所以模型辨识后还要经过考核.

一般地说，一个线性模型，参数确定了，得到线性方程

$$w_t-\hat{a}_1w_{t-1}-\cdots-\hat{a}_pw_{t-p}=n_t-\hat{b}_1n_{t-1}-\cdots-\hat{b}_qn_{t-q},$$

把样本数据 $w_1,w_2,\cdots,w_n$ 代进去，可以得到 $n_1,n_2,\cdots,n_n$，模型的考核就是要检查残量 $n_1,n_2,\cdots,n_n$ 是不是白噪声.

残量 n_t 的计算可按下列公式计算

$$n_t=w_t-\hat{a}_1w_{t-1}-\hat{a}_2w_{t-2}-\cdots-\hat{a}_pw_{t-p}+\hat{b}_1n_{t-1}+\hat{b}_2n_{t-2}+\cdots+\hat{b}_qn_{t-q}\qquad t=1,2,3,\cdots.$$

其中，$w_0, w_{-1}, \cdots, w_{1-p}, n_0, n_{-1}, \cdots, n_{1-q}$均为零

检验 n 个残量是不是白噪声的重要内容是检验 n 个残量是否为互不相关．

令残量的样本自协方差函数

$$\hat{r}_k(n,n)=\frac{1}{n}\sum_{t=1}^{n-k} n_t n_{t-k}$$

残量的样本自相关系数

$$\hat{\rho}_k(n,n)=\frac{\hat{r}_k(n,m)}{\hat{r}_0(n,n)} \qquad (k=1,2,\cdots,k)$$

其中，$\hat{r}_k(n,n)$，$\hat{\rho}_k(n,n)$中圆弧中第一个 n 表示样本长度，第二个 n 表示样本自协方差函数与样本自相关函数是对残量 n_t 讲的．k 值的选取为$\frac{n}{10}$左右．

计算 $Q_k=n\sum_{k-1}^{k}\hat{\rho}_k^2(n,n)$，采用 χ^2 检验．取置信率 ε，查表得 $x_{k,\varepsilon}^2$．若 $Q_k<\chi_{k,\varepsilon}^2$，则认为 n 个残量是互不相关的白噪声，即认为确定的线性模型是合适的．否则需要重新考虑模型的类别、阶数和参数．

§8.5　一些特定形式的模型

在许多实际问题中，由量测得到的随机序列多数并不平稳，而可能含有某种随机时间稳定增长或衰减的趋势，也可能含有随时间而周期式的变化起伏的趋势，这时不能采用平稳过程的模型，而需要用更一般的模型来描述

$$x_t=\mu_t+y_t$$

其中，μ_t 表示 x_t 中随时间变化的均值，它往往可用多项式、指数函数、正弦函数等来描述，而 y_t 表示零均值平稳过程，可以用 AR、MA 或 ARMA 模型拟合．

处理这类问题的方法可分为两类，一类方法是通过某些处理

方法剔除 μ_t 的变化趋势(例如 ARIMA 模型和乘积型季节模型)；另一类方法是具体求出 μ_t 的拟合形式(如混合回归模型).

一、ARIMA 模型

假如可以用多项式描述非平稳过程 x_t 中随时间变化的均值.

先通过一些具体例子来说明非平稳时间序列线性模型的形式.

如果非平稳时间序列 $\{z_t\}$ $(t=1,2,\cdots)$ 可以表示成

$$z_t=a+b_t+x_t$$

$\{x_t\}$ 是均值为零的平稳时间序列. 从图 8-6 可见,它的图像是绕直线 $y=a+bt$ 上下作均匀波动的一条折线.

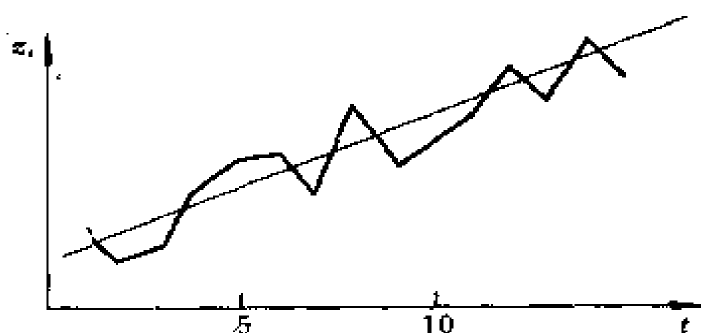


图 8-6　非平稳时间序列 $\{z_t\}$(1)

对 z_t 作差分运算,则 $\nabla z_t=z_t-z_{t-1}=a+bt+x_t-[a+b(t-1)+x_{t-1}]=b+x_t-x_{t-1}$ $(t\geqslant 2)$,数学上可以证明 $\{b+x_t-x_{t-1}\}$ 是平稳时间序列. 由平稳时间序列的线性模型可知,∇z_t 的线性模型也有自回归模型、滑动平均模型、混合模型三种. 写成一般形式(式中允许 $p=0$ 或 $q=0$)

$$(\nabla z_t-\mu)+\varphi_1(\nabla z_{t-1}-\mu)+\cdots+\varphi_p(\nabla z_{t-p}-\mu)$$
$$=\varepsilon_t+\theta_1\varepsilon_{t-1}+\cdots+\theta_q\varepsilon_{t-q}$$

其中 $\mu=E(\nabla z_t)$,令 $\theta_0=\mu(1+\varphi_1+\cdots+\varphi_p)$,则

$$(z_t-z_{t-1})+\varphi_1(z_{t-1}-z_{t-2})+\cdots+\varphi_p(z_{t-p}-z_{t-p-1})$$
$$=\theta_0+\varepsilon_t+\theta_1\varepsilon_{t-1}+\cdots+\theta_q\varepsilon_{t-q}$$

这就是非平稳时间序列$\{z_t\}$的线性模型.

再如非平稳时间序列$\{z_t\}$可以表示成

$$z_t = b_0 + b_1 t + b_2 t^2 + x_t$$

其中$\{x_t\}$是均值为零的平稳时间序列,它的图形是绕一条二次曲线上下均匀摆动的折线,如图 8-7 所示.

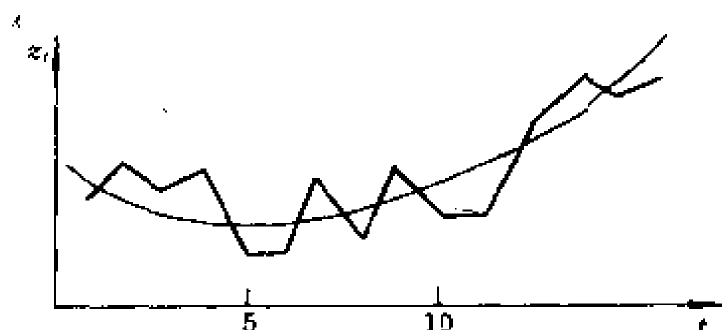


图 8-7 非平稳时间序列$\{z_t\}$(2)

此处作一阶差分不可能得到平稳时间序列. 作二阶差分$\nabla^2 z_t = z_t - 2z_{t-1} + z_{t-2}$ $(t \geqslant 3)$

可得
$$\nabla^2 z_t = 2b_2 + x_t - 2x_{t-1} + x_{t-2}$$

可以证明$\{\nabla^2 z_t\}$是平稳时间序列,其线性模型是

$$(\nabla^2 z_t - \mu) + \varphi_1(\nabla^2 z_{t-1} - \mu) + \cdots + \varphi_p(\nabla^2 z_{t-p} - \mu)$$
$$= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

其中,$\mu = E(\nabla^2 z_t)$,或写成

$$(z_t - 2z_{t-1} + z_{t-2}) + \varphi_1(z_{t-1} - 2z_{t-2} + z_{t-3}) + \cdots$$
$$+ \varphi_p(z_{t-p} - 2z_{t-p-1} + z_{t-p-2})$$
$$= \theta_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \qquad (t \geqslant 3)$$

其中$\theta_0 = \mu(1 + \varphi_1 + \cdots + \varphi_p)$,这就是非平稳时间序列$\{z_t\}$的线性模型.

一般地说,如果非平稳时间序列$\{z_t\}$可表示为$z_t = b_0 + b_1 t + \cdots + b_d t^d + x(t)$ $(d \geqslant 0)$

$\{x_t\}$是均值为零的平稳时间序列. 作 d 阶差分

$$\nabla^d z_t = z_t - d z_{t-1} + \frac{d(d-1)}{2} z_{t-2} - \cdots + (-1)^d z_{t-d} \quad (t \geqslant d+1)$$

时间序列$\{\nabla^d z_t\}(t\geqslant d+1)$是平稳的．平稳时间序列$\{\nabla^d z_t\}$的线性模型是

$$
\begin{aligned}
&(\nabla^d z_t-\mu)+\varphi_1(\nabla^d z_{t-1}-\mu)+\cdots+\varphi_p(\nabla^d z_{t-p}-\mu)\\
&\quad=\varepsilon_t+\theta_1\varepsilon_{t-1}+\cdots+\theta_q\varepsilon_{t-q}\quad(t\geqslant d+1)
\end{aligned}
$$

其中$\mu=E(\nabla^d z_t)$. 这种模型称为求和模型,它的阶数为(p,d,q),求和模型简记为 ARIMA(p,d,q). 实际问题遇到的差分阶数 d 一般不超过二阶．

对一个非平稳时间序列进行一系列观测,得到一个样本 z_1, $z_2,\cdots,z_n$,如何确定它的线性模型呢? 这里有两种方法．

方法一,先把样本数据 $z_t, t=1,2,\cdots,n$ 画成图,看它围绕 n 次代数曲线作均匀摆动,于是对 z_t 作 n 次差分．记差分次数为 d. 然后检验$\nabla^d z_t$ 是否属平稳时间序列．这可用 $n-d$ 个$\nabla^2 z_t$ 的数据作出样本自相关函数和样本偏相关函数,分别作出点图,如果两个点图是拖尾、截尾的,那么$\nabla^d z_t$ 一定是平稳的．最后可用模型辨识和参数估计的方法确定模型．这种方法的缺点是在从 z_t 的点图中不易看出它到底绕几次曲线作均匀摆动,因此对 z_t 的差分阶数 d 不易正确定出．

方法二,对于 $d=0$,计算$\nabla^d z_t$ 的样本自相关函数和样本偏相关函数,分别作出点图,如果两个点图分别是拖尾和截尾的,则表明 z_t 是平稳时间序列,进一步确定其模型,否则,$\nabla^d z_t$ 不是平稳的,令 $d=d+1$,重复上述步骤,直到$\nabla^d z_t$ 是平稳时间序列,作出线性模型为正．关于 z_t 的各阶差分的计算可采用表 8-6 格式．

表 8-6　差分计算表

原始数据	z_1	z_2	z_3	z_4	…	z_n
一阶差分		$\nabla z_2=z_2-z_1$	$\nabla z_3=z_3-z_2$	$\nabla z_4=z_4-z_3$	…	$\nabla z_n=z_n-z_{n-1}$
二阶差分			$\nabla^2 z_3=\nabla z_3-\nabla z_2$	$\nabla^2 z_4=\nabla z_4-\nabla z_3$	…	$\nabla^2 z_n=\nabla z_n-\nabla z_{n-1}$
三阶差分				$\nabla^3 z_4=\nabla^2 z_4-\nabla^2 z_3$	…	$\nabla^3 z_n=\nabla^2 z_n-\nabla^2 z_{n-1}$
…	…			…	…	…

例 8-5 对某化学反应过程的 226 个温度记录数据 z_t，计算 $\hat{\alpha}_{kk}$截尾，但 $\hat{\rho}_k$既不截尾，亦不拖尾，可以认为 z_t 是非平稳的．

对 z_t 取一阶差分，得到 225 个∇z_t 的数据，计算得 $\hat{\rho}_k$与 $\hat{\theta}_{kk}$的数值见表 8-7．

表 8-7　∇z_t 的 $\hat{\rho}_k$与 $\hat{\alpha}_{kk}$数值表

k	1	2	3	4	5	6	7	8
$\hat{\rho}_k$	0.81	0.66	0.53	0.45	0.39	0.33	0.27	0.19
$\hat{\alpha}_{kk}$	0.81	0.01	−0.01	0.05	0.03	0.03	−0.05	0.02
k	9	10	11	12	13	14	15	
$\hat{\rho}_k$	0.14	0.15	0.10	0.10	0.08	0.08	0.08	
$\hat{\alpha}_{kk}$	0.80	0.02	−0.02	−0.08	0.08	0.13	−0.15	

由表可见，$\hat{\rho}_k$拖尾，$\hat{\alpha}_{kk}$截尾在 $k=1$ 处，所以∇z_t 属于一阶自回归模型．令 $w_t=\nabla z_t-\overline{\nabla z_t}$，其中，

$$\overline{\nabla z_t}=\sum_{t=2}^{226}\nabla z_t\Big/225$$

于是关于 w_t 的模型方程的形式是

$$w_t+\varphi_1 w_{t-1}=\varepsilon_t$$

用 z_t 表示时，可写成

$$(\nabla z_t-\overline{\nabla z})+\varphi_1(\nabla z_{t-1}-\overline{\nabla z})=\varepsilon_t$$

或
$$(z_t-z_{t-1}-\overline{\nabla z})+\varphi_1(z_{t-1}-z_{t-2}-\overline{\nabla z})=\varepsilon_t.$$

如果对 z_t 取二阶差分，得到 224 个$\nabla^2 z_t$的数据，计算得 $\hat{\rho}_k$和 $\hat{\alpha}_{kk}$的数据见表（表 8-8）．

由表可见，$\hat{\rho}_k$与 $\hat{\alpha}_{kk}$均截尾截在 $k=0$ 处，所以$\nabla^2 z_t$ 属于零阶自回归模型，记

$$w_t \triangleq \nabla^2 z_t - \overline{\nabla^2 z}, \overline{\nabla^2 z} = \sum_{t=3}^{26} \nabla^2 z_t / 22\varphi$$

则 $w_t = \epsilon_t$，或 $z_t - 2z_{t-1} + z_{t-2} - \overline{\nabla^2 z} = \epsilon_t$.

表 8-8 $\nabla^2 z_t$ 的 $\hat{\rho}_k$ 与 $\hat{a}_{kk}$ 数值表

k	1	2	3	4	5	6	7	8
$\hat{\rho}_k$	−0.08	−0.07	−0.12	−0.06	0.019	−0.02	0.05	−0.05
$\hat{a}_{kk}$	−0.08	−0.07	−0.14	−0.10	−0.02	−0.05	0.02	−0.06
k	9	10	11	12	13	14	15	
$\hat{\rho}_k$	−0.13	0.13	−0.13	0.08	−0.08	0.03	−0.01	
$\hat{a}_{kk}$	−0.15	0.10	−0.15	0.02	−0.09	−0.01	−0.04	

二、乘积型季节性模型

在实际中，经常遇到一些依时间而周期性变化的观察序列，例如各种气象观察资料、各种经济信息、各种天文、地震观测数据等等，形象地把这类变化规律称为"季节"性变化．在许多问题中，往往可以从直观背景及物理变化规律得到数据序列的周期．很自然地想到，如果把某一时刻的观察值与下一周期相应时刻的观察值相减，就可能将周期性变化消除掉．例如考察月平均气温的时间序列$\{x_t\}$，显然数据应有年周期(即周期 $s=12$ 月)，若把每个月的气温与去年同月气温相减，得到的新序列便是在各月平均气温附近的波动值，因而相当接近于平稳序列．不难验证，一阶季节差分后，可消除 x_t 中周期为 s 的正弦分量 $\sin(2\pi t/s)$；二阶季节差分后，x_t 中含有 $t\sin(2\pi t/s)$的分量也消除了．

若定义差分算子$\nabla = 1-B$ 和季节差分算子为

$\nabla_s = 1-B^s, \nabla_s^D = (1-B^s)^D$,

$U(B^s) = 1 + u_1 B^s + u_2 B^{2s} + \cdots + u_p B^{ps}$,

$$V(B^s)=1+v_1B^s+v_2B^{2s}+\cdots+v_qB^{qs}$$

则原序列经▽阶季节差分后得到的序列$\{x_t'\}$有

$$x_t'=\nabla_s^D x_t$$

就周期点上来考察是平稳的，因而可建立关于周期s的时间序列模型(称为季节性模型)：

$$U(B^s)x_t'=V(B^s)E_t \tag{1}$$

假定(1)式的E_t不一定是白噪声，例如是另一个ARIMA(n,d,m)序列

$$\varphi(B)\nabla^d E_t=\theta(B)\varepsilon_t$$

其中ε_t是白噪声，则得到乘积型季节模型

$$\varphi(B)U(B^s)\nabla^d D_s^d x_t=\theta(B)v(B^s)\varepsilon_t$$

可将(1)式展开成通常的ARIMA模型形式．例如将

$$(1-B)x_t=(1+b_1B)(1+v_1B^s)\varepsilon_t$$

展开成

$$\begin{aligned}(1-B)x_t&=(1+b_1B+v_1B^s+b_1v_1B^{s-1})\varepsilon_t\\&=\sum_{j=0}^{s+1}b'_jB^j\varepsilon_t.\end{aligned}$$

这是一个$(0,1,s+1)$阶ARIMA模型，但参数$b'_1,b'_2,\cdots,b'_{s-1}$有约束关系$b'_1=b,b'_2=\cdots=b'_{s-1}=0,b'_s=v_1,b'_{s+1}=b_1v_1$，尽管模型的阶数$(s+1)$可能很高，但除$b'_1,b'_s,b'_{s-1}$以外，其他参数都是零，因此是疏系数模型，而且参数$b'_{s+1}=b_1v$，实际上只有$b'_1$和$b'_s$是自由参数$(b_0=b'_0\equiv1)$．

三、组合模型

通常称时序模型$x_t=u_t+y_t$中的u_t为序列的确定性部分，而称y_t为零均值的平稳随机部分，希望不仅要将趋势性和周期性分量分离出来，而且要给出u_t的具体表达式．这种既有确定性又有随机性部分组合而成的模型．在描述某些类型的非平稳过程时，常常有较好的效果．

建立组合模型，通常是先用最小二乘方法按某类函数拟合数据序列的确定部分，从低阶开始，逐渐增加阶数，直到模型无明显改进为止．然后对消除了确定趋势的残量序列建立适宜的 ARMA(n,m)模型．最后用前述得到的两部分参数估值作为初值，对确定性部分和 ARMA 部分的所有参数，同时用非线性最小二乘方法重新估计，得到组合模型的最终估计．

常用的组合模型有：

1° 多项式趋势

$$x_t=\sum_{j=0}^{r}b_jt^j+Y_t$$

2° 指数趋势

$$x_t=\sum_{j=1}^{r}A_je^{k_jt}+Y_t$$

3° 周期趋势

$$x_t=\sum_{j=1}^{r}A_je^{k_jt}+\sum_{j=1}^{r}B_je^{l_jt}\sin(jwt+\varphi_j)+Y_t$$

为计算方便，常改写成

$$x_t=\sum_{j=1}^{r}A_je^{k_jt}+\sum_{j=1}^{s}B_je^{l_jt}[c_j\sin(jwt)+\sqrt{1-c_j}\cos(jwt)]+Y_t$$

4° 一般形式

组合模型可用如下形式统一描述：

$$x_t=\sum_{j=0}^{M}A_je^{k_jt}+Y_t$$

如果 A_j 全为零，则 $x_t=Y_t$，即为普通 ARMA(n,m)平稳序列．若仅有一个 A_1 非零，且与其相应的 k_j 为零，则$\{x_t\}$为均值为常数的平稳序列．若 m 个 k_j 中只存在值很小的实数时，将指数函数作 Taylor 展开，则只有前几项较大，$\{x_t\}$表现为具有多项式趋势．当 k_j 为较大的正负实数时，模型显示出具有指数增长或衰减趋势．当复指数 k_j 具有负实部时，$\{x_t\}$有阻尼的正弦余弦趋势．当复指数 k_j 实部为零时，就是带随机干扰的正余弦趋向，实部为正时，则

是振幅不断增大的正余弦趋向．

四、门限自回归模型

门限自回归模型(Threshold Autoregressive Model)用来解决一类非线性问题，其基本思想是：把非线性模型按照某一变元的不同取值范围，采用若干个线性模型来描述．它将微分方程中极限环的概念引入非线性随机系统，可以有效地描述具有周期规律的过程，由于门限的控制作用，保证了模型的稳定性．这类模型还可以做为突变现象的一种描述手段．

通常采用的门限回归模型有三种定义，设$\{z_t\}$是因变元观察序列，$\{x_{it}, 1\leqslant i\leqslant s\}$是自变元观察序列：

$$z_t=\begin{cases} a_0^{(1)}+a_1^{(1)}x_{1t}+\cdots+a_s^{(1)}x_st+\varepsilon_t^{(1)}(\text{当 } x_{l,t-d}\leqslant \bar{x}_1) \\ a_0^{(2)}+a_1^{(2)}x_{1t}+\cdots+a_s^{(2)}x_st+\varepsilon_t^{(2)}(\text{当 } \bar{x}_1<x_{l,t-d}\leqslant \bar{x}_2) \\ \cdots \\ a_1^{(r)}+a_1^{(r)}x_{1t}+\cdots+a_s^{(r)}x_{st}+\varepsilon_t^{(r)}(\text{当 } \bar{x}_{r-1}<x_{l,t-d}<+\infty) \end{cases}$$

其中自变元序列$\{x_{lt}\}$称为门限变元，$\bar{x}_i, i=1,\cdots,r-1$ 为门限值，d为门限元滞后量．

设$\{z_t\}$是单变量观察序列，

$$z_t=\begin{cases} \beta_0^{(1)}+\beta_1^{(1)}z_{t-1}+\cdots+\beta_{p_1}^{(1)}z_{t-p_1}+\varepsilon_t^{(1)} & (\text{当 } z_{t-d}\leqslant \bar{z}_1) \\ \beta_0^{(2)}+\beta_1^{(2)}z_{t-1}+\cdots+\beta_{p_2}^{(2)}z_{t-p_2}+\varepsilon_t^{(2)} & (\text{当 } \bar{z}_1<z_{t-d}\leqslant \bar{z}_2) \\ \cdots \\ \beta_0^{(r)}+\beta_1^{(r)}z_{t-1}+\cdots+\beta_{p_r}^{(r)}z_{t-p_r}+\varepsilon_t^{(r)} & (\text{当 } \bar{z}_{r-1}<z_{t-d}) \end{cases}$$

其中 $\bar{z}_1,\bar{z}_2,\cdots,\bar{z}_{r-1}$为门限值，$d$ 是门限滞后量．

$$
z_t=\begin{cases}
\beta_0^{(1)}+\beta_1^{(1)}z_{t-1}+\cdots+\beta_{q_1}^{(1)}z_{t-q_1}+\sum_{j=1}^{s}\sum_{i=0}^{p_j^{(1)}}a_{ji}^{(1)}x_{j,t-i} \\
\qquad\qquad(\text{当 } x_{t,t-d}\leqslant\bar{x}_1) \\
\beta_0^{(2)}+\beta_1^{(2)}z_{t-1}+\cdots+\beta_{q_2}^{(2)}z_{t-q_2}+\sum_{j=1}^{s}\sum_{i=0}^{p_j^{(2)}}a_{ji}^{(2)}x_{j,t-i} \\
\qquad\qquad(\text{当 } \bar{x}_1<x_{t,t-d}\leqslant\bar{x}_2) \\
\cdots \\
\beta_0^{(r)}+\beta_1^{(r)}z_{t-1}+\cdots+\beta_{q_r}^{(r)}z_{t-q_r}+\sum_{j=1}^{s}\sum_{i=0}^{p_j^{(r)}}a_{ji}^{(r)}x_{j,t-i} \\
\qquad\qquad(\text{当 } \bar{x}_{r-1}<x_{t,t-d})
\end{cases}
$$

其中 $\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_{r-1}$ 为门限值，$\{x_{tt},1\leqslant t\leqslant N\}$ 为门限变元，d 是门限滞后．显然定义 1 和 2 都是定义 3 的特例．

门限自回归模型是具有较广泛意义的一种非线性模型，例如考察一阶非线性自回归模型

$$x_t=f(x_{t-1})+\varepsilon_t$$

其中 f 是连续函数，且在区间$[a,b]$上为一致连续．对任何 $x\in[a,b]$，利用近似展开，存在关系式：

$$f(x)\doteq f(\bar{x}_{j-1})+\frac{f(\bar{x}_j)-f(\bar{x}_{j-1})}{\bar{x}_j-\bar{x}_{j-1}}(x-\bar{x}_{j-1})=a_{j-1}+a_jx$$

其中$a_j\triangleq\dfrac{f(\bar{x}_j)-f(\bar{x}_{j-1})}{\bar{x}_j-\bar{x}_{j-1}}$，$a_{j-1}\triangleq f(\bar{x}_{j-1})-a_j\bar{x}_{j-1}$，

$a=\bar{x}_0<\bar{x}_1<\cdots<\bar{x}_r=b$ 是$[a,b]$上的一组分割，分割的粗细由逼近精度决定．在给定精度下，一阶非线性回归模型可以由下式近似：

$$x_t=a_{j-1}+a_jx_{t-1}+\varepsilon_t\quad(\bar{x}_{j-1}\leqslant x_{t-1}<\bar{x}_j)$$

这正是一阶门限自回归模型．可见一阶非线性回归模型可由一阶门限自回归模型近似．

类似地，对于一般的 k 阶非线性自回归模型 $x_t=f(x_{t-1},\cdots,x_{t-k})+\varepsilon_t$ 也可用高阶的门限自回归模型逼近．这说明了门限自回

归模型的广泛意义．

可以证明线性随机差分方程的结构和性质与对应的齐次差分方程有密切关系，若齐次差分方程的解当 $t\to\infty$ 时趋于零，则相应自回归模型是平稳的．所以可以利用齐次差分方程了解门限自回归模型的特性．令

$$f(x_{t-1},x_{t-2},\cdots,x_{t-p})=\begin{cases}\beta_0^{(1)}+\beta_1^{(1)}x_{t-1}+\cdots+\beta_{p_1}^{(1)}x_{t-p_1} & (\text{当 } x_{t-d}\leqslant\bar{x}_1)\\ \beta_0^{(2)}+\beta_1^{(2)}x_{t-1}+\cdots+\beta_{p_2}^{(2)}x_{t-p_2} & (\text{当 } \bar{x}_1<x_{t-d}\leqslant\bar{x}_2)\\ \cdots & \\ \beta_0^{(r)}+\beta_1^{(r)}x_{t-1}+\cdots+\beta_{p_r}^{(r)}x_{t-p_r} & (\text{当 } \bar{x}_{r-1}<x_{t-d}<+\infty)\end{cases}$$

其中 $P=\max(P_1,P_2,\cdots,P_r)$．则与门限自回归模型相应的齐次差分方程为

$$x_t=f(x_{t-1},x_{t-2},\cdots,x_{t-p}).\tag{2}$$

上式的解具有如下特性：

1° 若从某一初始值 $x_P,x_{P-1},\cdots,x_1$ 出发，(2)式给出的序列 x_t 满足 $\lim\limits_{t\to\infty}x_t=x$，则称 x 为(2)式的极限点．

2° 若从某组初始值 $x_P,x_{P-1},\cdots,x_1$ 出发，x_t 渐近地与某一周期为 T 的周期序列相吻合，即

$$\lim_{t\to\infty}\sum_{i=1}^{T}(x_{t-i}-c_{t-i})^2=0$$

则称 $\{c_t\}$ 是（＊）式以 T 为周期的极限环．T 是使 C_t 满足下式的最小周期：

$$C_{t+T}=C_t \qquad t=1,2,3,\cdots$$

若从任何 $(x_p,x_{p-1},\cdots,x_1)$ 的邻域出发，$\{x_t\}$ 都与 $\{C_t\}$ 渐近吻合，则称它为稳定极限环．

例如，考虑一阶齐次差分方程

$$x_t=\begin{cases}4x_{t-1}, & \text{当 } |x_{t-1}|\leqslant\frac{1}{2}\\ \frac{1}{2}x_{t-1}, & \text{当 } |x_{t-1}|>\frac{1}{2}\end{cases}$$

取初值 $x_1=4$，则 $x_2=2, x_3=1, x_4=\frac{1}{2}, x_5=2, x_6=1, x_7=\frac{1}{2},\cdots$，即 $\{x_t\}$ 有极限环 $\{\frac{1}{2},2,1,\frac{1}{2},1,2,1,\cdots\}$. 若取初值 $x_1=5$，则 $x_2=\frac{5}{2}, x_3=\frac{5}{4}, x_4=\frac{5}{8}, x_5=\frac{5}{16}, x_6=\frac{5}{4}, x_7=\frac{5}{8}, x_8=\frac{5}{16},\cdots$，$\{x_t\}$ 具有极限环 $\{\frac{5}{8},\frac{5}{16},\frac{5}{4},\frac{5}{8},\frac{5}{16},\frac{5}{4},\cdots\}$. 然而对于一般的差分方程很难详尽举出它所有的极限点和极限环．把微分方程极限环的概念引入随机的门限自回归模型中，进一步探讨了门限自回归模型的特性，及适合这类模型的周期序列与极限环的关系．

由于门限回归（自回归）模型是按照门限元不同的取值范围而采取不同的拟合模型，首先要把门限变元 $\{x_{it}, t=1,2,\cdots,N\}$ 按照从小到大的顺序进行排列，而门限值 $\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_{r-1}$ 及 r 的选取，一般是通过经验分布确定．如取 $\{x_{it}, l, t\leqslant N\}$ 作为门限变量，那么我们就要计算出它们的经验概率密度，进而求出经验分布 $F(x_{t,i})$，为了使每两个门限值之间出现的观察数据不至过少，一般选门限值为 $\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_9$，使 $F(\bar{x}_1)=0.1, F(\bar{x}_2)=0.2,\cdots,F(\bar{x}_9)=0.9$. 若从 $\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_9$ 中挑选 $r-1$ 个最优门限值，就是寻求使门限回归模型的 AIC 函数值为最小的那组个数为 $r-1$ 的门限值．实际上一般取 $r=2,3$ 或 4（即门限分点数为 1，2 或 3）. 对不同的 r 值再进行比较以确定最佳模型的门限值个数．

例 8-6 1821 年至 1934 年期间捕捉的加拿大山猫数的年度记录（以 10 为底取对数），共 114 个观察值，如表 8-8 和图 8-8 所示，试建立加拿大山猫数量的数学模型，并预测变化趋势．

上述数据描述了一种生态自然循环现象．在一个给定地区内的野兔和山猫的总数具有某种循环或振荡变化的形式，野兔是山

猫的主要食物，当野兔数量很大时，山猫的数量增多，然而不久随着山猫成倍增长，由于过量捕食的结果，野兔数量下降，而这又引起部分山猫饿死，从而山猫数量减少，这就完成一个循环．如果野兔的数量比供山猫捕食所需野兔的数量大，则山猫数量又要增加．从观察数据可以看出其变化有近十年的明显周期．由于数据变化起伏很大，已事先对数据取对数．

表 8-9　1921 年—1934 年加拿大山猫数据

269	321	585	871	1475	2821	3928	5943	4950	2577	523	98
184	279	409	2285	2685	3409	1824	409	151	45	68	213
546	1033	2129	2536	957	361	377	225	360	731	1638	2725
2871	2119	684	299	236	245	552	1623	3311	6721	4254	687
255	473	358	784	1594	1676	2251	1426	756	299	201	229
469	736	2042	2811	4431	2511	389	73	39	49	59	188
377	1292	4031	3495	537	105	153	387	758	1307	3465	6991
6313	3794	1836	345	382	808	1388	2713	3800	309	2985	3790
674	71	80	108	229	399	1132	2432	3575	2935	1537	529
485	662	1000	1520	2657	3396						

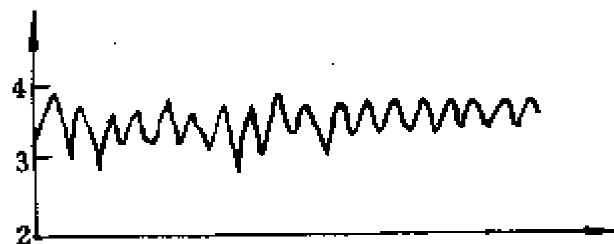


图 8-8　山猫数据及预报曲线

门限值取为 $\bar{z}=3.1163$，门限滞后量 $d=2$，得到的门限自回归模型是

$$z_t=\begin{cases}0.5239+1.0359z_{t-1}-0.1756z_{t-2}+0.1753z_{t-3}-0.4339z_{t-4} \\ \quad +0.3457z_{t-5}-0.3032z_{t-6}+0.2162z_{t-7} \\ \quad +0.0043z_{t-8}+\varepsilon_t^{(1)} \qquad (\text{当 } z_{t-2}\leqslant 3.1163) \\ 2.6559+1.4246z_{t-1}-1.1618z_{t-2}-0.1094z_{t-3}+\varepsilon_t^{(2)} \\ \qquad\qquad (\text{当 } z_{t-2}>3.1163)\end{cases}$$

其中 $\mathrm{Var}[\varepsilon_t^{(1)}]=0.0255, \mathrm{Var}[\varepsilon_t^{(2)}]=0.0516$.

从图 8-7 中可以看到，最终预报曲线是一个周期 9 年的非对称函数，这表明该模型存在以 9 年为周期的极限环，上升段和下降段长度分别为 6 年和 3 年，它的极限环为(2.6226，2.8945，3.2525，3.4601，3.4257，3.2281，2.9793，2.7884，2.6639，…)滞后量 $d=2$ 表明山猫数量的增减与捕获量之间存在着滞后两年的依赖关系.

（该问题的解析模型可参考第五章中弱肉强食模型一节）.

§8.6 非线性系统模型参数的估计

在实际问题中遇到的系统，大多数是非线性系统．当系统的非线性环节很近似于线性特性时，用线性理论来分析和设计，在工程上是允许的，但对一般的非线性系统则不行．非线性系统的辨识也有稳态系统模型和动态系统模型辨识之分．这里着重讨论非线性动态系统的辨识．

非线性动态系统模型和线性动态系统的描述方法类似，也用微分方程表达式和状态空间表达式．一般非线性系统可用 Volterra 级数来表示其输出

$$y(t)=\int_0^t g_1(\tau)u(t-\tau)\mathrm{d}\tau+\int_0^t\int_0^t g_2(\tau_1,\tau_2)u(t-\tau_1)u(t-\tau_2)\mathrm{d}\tau_1\mathrm{d}\tau_2 + \int_0^t\int_0^t\int_0^t g_3(\cdots)\cdots+\cdots$$

其中 $g_i(\tau_1,\tau_2,\cdots,\tau_i)$ 称为 Volterra 级数的核．这个级数是由 $y=f(u_1,u_2,\cdots,u_n)$ 的多维泰勒级数展开和类似的叠加原理并取极限

而推得．可以看成是线性系统脉冲响应函数卷积表达式的推广．

Volterra 级数表示法，有两种特殊情况研究得较多，这就是哈默斯坦（Hammerstein）模型和维纳（Wiener）模型．

图 8-9 所示的非线性系统模型为 Hammerstein 模型，简称 H 模型．图中 HL 表示单值非线性特性，其输出为

$$u(t)=r_1 l(t)+r_2 l^2(t)+\cdots+r_n l^n(t)$$

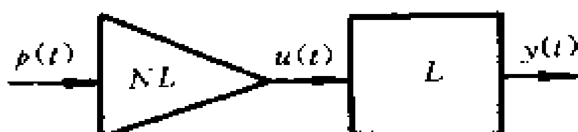


图 8-9　非线性系统的 Hammerstein 模型

图 8-9 中线性部分特性为

$$y(t)=\int_0^T g(\tau)u(t-\tau)\mathrm{d}\tau$$

其中，T 是系统的调整时间，即当 $\tau>T$ 时，$g(\tau)=0$. 整个 H 模型的数学表达式为

$$y(t)=\int_0^t g(\tau)[r_1 l(t-\tau)+r_2 l^2(t-\tau)+\cdots+r_n l^n(\tau)]\mathrm{d}\tau$$

离散的广义 H 模型，可以写成

$$y(k)=k_0+\frac{B_1(z^{-1})}{A(z^{-1})}u(k-d)+\cdots+\frac{B_p(z^{-1})}{A(z^{-1})}u^p(k-d)$$
$$+\frac{c(z^{-1})}{A(z^{-1})}\varepsilon(k)\quad k=1,2,\cdots$$

其中，

$$\begin{cases}A(z^{-1})=1+a_1 z^{-1}+\cdots+a_n z^{-n}\\ B_1(z^{-1})=b_{10}+b_{11}z^{-1}+\cdots+b_{1n}z^{-n}\\ \cdots\\ B_p(z^{-1})=b_{p0}-b_{p1}z^{-1}+\cdots+b_{pn}z^{-1}\\ C(z^{-1})=1+c_1 z^{-1}+\cdots+c_n z^{-n}\end{cases}$$

k_0 是静态增益，d 是过程的滞后步数，z^{-1}是单位后移算子．

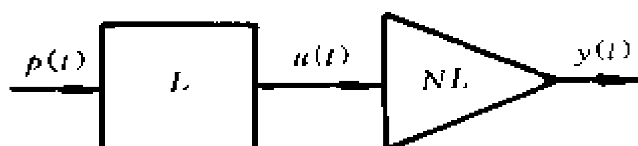


图 8-10　非线性系统的 Wiener 模型

图 8-10 所示为维纳(Wiener)模型，简称 W 模型．W 模型可以用下列数学表达式表示：

$$y(t)=r_1u(t)+r_2u^2(t)+\cdots+r_nu^n(t),$$

其中，$u(t)=\int_0^T g(\tau)l(t-\tau)\mathrm{d}\tau$

以 Harmmerstein 模型为例，设线性系统用差分方程模型描述，阶为 n，即

$$A(z^{-1})y(k)=B(z^{-1})u(k)$$

其中　　$A(z^{-1})=1+a_1z^{-1}+\cdots+a_nz^{-n}$

$$B(z^{-1})=b_0+b_1z^{-1}+\cdots+b_nz^{-n}$$

若线性系统是稳定的，输出端的噪声为零均值的随机变量，问题归结为：设 n 及 p 已知，如何利用观测序列 $l(k)$ 和 $y(k)$，估计参数 a_i，b_i 和 r_i．考虑 Harmmerstein 模型

$$A(z^{-1})y(k)=B(z^{-1})[l(k)+\sum_{i=2}^{p}r_il^i(k)]+\varepsilon(k) \qquad (3)$$

其中，不失一般性，r_1 已规格化，取为 1.

由于出现了未知参数 b_i 和 r_i 的乘积，我们面临着一个非线性参数的估计问题，可以分两步估计．第一步先估计出乘积项 r_ib_i，然后再从乘积项中把 r_i 和 b_i 分离开．

首先定义多项式 $S(z^{-1})$

$$S(z^{-1})=\sum_{i=2}^{p}r_iB(z^{-1})=s_0+s_1z^{-1}+\cdots+s_nz^{-n}$$

其中 $s_j=b_j\sum_{i=2}^{p}r_i \qquad j=0,1,2,\cdots,n.$

化简(1)式得到

$$A(z^{-1})y(k)=B(z^{-1})l(k)+S(z^{-1})\sum_{i=2}^{p}l^{i}(k)+\varepsilon(k) \quad (4)$$

显然 $A(z^{-1})$、$B(z^{-1})$和 $S(z^{-1})$和测量数据 $y(k)$、$l(k)$成线性组合关系. 如果 $\varepsilon(k)$是白噪声,利用最小 2 乘估计,可得

$\hat{a}_i(i=1,2,\cdots,n)$,$\hat{b}_i(i=0,1,\cdots,n)$和 $\hat{s}_i(i=0,1,\cdots,n)$

的无偏估计. 代入(2)式,

$$\begin{cases}\hat{S}_0=\hat{b}_0r_2+\hat{b}_0r_3+\cdots+\hat{b}_0r_p\\ \hat{S}_1=\hat{b}_1r_2+\hat{b}_1r_3+\cdots+\hat{b}_1r_p\\ \cdots\\ \hat{S}_n=\hat{b}_nr_2+\hat{b}_nr_3+\cdots+\hat{b}_nr_p\end{cases}$$

再利用最小二乘法. 可得

$$\hat{r}_i,\quad i=2,3,\cdots,p$$

的最小二乘估计.

假如关心的是系统稳定状态输入输出的关系,利用回归分析方法处理稳态模型,如第七章讨论的. 对于高阶动态系统,或当过程的运动规律比较复杂,或者过程运动很慢,常常用较为简单的静态模型来描述,若考虑的过程沿某一直线或曲线表现为零均值,而且关心的是过程由一个稳定状态变到另一个稳定状态的过渡过程,即更关心系统的动态特性,则采用系统辨识方法.

系统辨识的方法很多,这里只介绍了参数模型中的时间序列分析方法,而没有涉及其他的重要方法,例如传递函数的辨识方法及其他非参数模型. 利用时间序列分析,可以进行系统的预报、控制和诊断,此处仅限于预报,因此讨论了模型的参数估计、结构辨识和模型辨识,而略去了在控制中有重要作用的状态估计与滤波以及诊断中的谱分析理论.

在时间序列分析中介绍了三种最重要的线性模型,非线性模型特别是门限自回归以及非平稳时间序列中的季节模型和组合模型.

最小二乘参数估计主要是用于输入输出形式的差分模型的参数估计上，对于线性系统的模型类的向量矩阵表达式为 $Y=\varphi\theta+\varepsilon$，其正规方程组为 $\hat{\varphi}^T\varphi\hat{\theta}=\varphi^TY$，参数解的表达式为 $\hat{\theta}=(\hat{\varphi}^T\varphi)^{-1}\varphi^TY$. 这是线性静态系统和动态系统参数估计上的共同点．然而线性静态系统的模型中，φ 与 ε 不相关，ε 也是零均值不相关系列，因此最小二乘参数估计是无偏的和一致的．线性动态系统的模型中，φ 与 ε 相关，一般情况下 ε 也是相关序列，所以其参数估计是有偏的（工程上允许）和非一致的，需要进一步改进．改进的方法有：广义最小二乘法、辅助变量法和增广矩阵法，以及将一般最小二乘法与其他方法相结合的方法——最小二乘两步法和随机逼近算法.

在非线性系统中，模型结构的建立主要通过机理分析的方法，而通过这种方法所得到的模型结构相当大的部分是属于稳态非线性模型，因此稳态非线性模型的参数估计在非线性系统的辨识中占主要地位，而其参数估计问题往往又归结为最优化方法，通常采用非线性最小二乘法．系统的非线性，使模型极其丰富也造成建模的困难. 门限自回归模型是具有较广泛意义的一种非线性模型，其思想方法类似于实变函数中的勒贝格积分（Lebesgue）.

系统辨识中目前引人关注的除非线性系统辨识外，还有神经网络用于系统辨识，自适应建模等等．所谓“自适应”是模型在某种程度上能实时地根据量测数据（输入）和估计结果（输出），自行调整模型参数，并随着数据的陆续到来，通过递推算法自动地对模型加以修正，使其接近某种最佳值，即便在尚不完全掌握序列特性的情况下也能得到满意的模型．递推算法主要基于维纳（Wiener）滤波理论的方法、基于卡尔曼（Kalman）滤波理论的方法以及最小二乘的递推算法．

系统的可辨识性在数学模型中占有重要地位，尤其在闭环系统和分布参数系统中．系统辨识是属于微分方程的反问题范畴，所谓“反问题”，指的是常常可以根据事物的物理、化学及生物变化，推导出该物理问题满足一个微分方程及其界面条件，但方程中

的某些系数未知，对偏微分方程而言，还可能对场存在的范围或者界面条件未知，需要利用场的某些信息反求这些未知数．反问题研究的主要内容是问题的适定性(即要求确定合理的充分条件，使得该反问题在条件限制的范围内有解、或者解唯一，或者解连续依赖于量测数据)，以及反问题的数值解法．

习　题

1. 有 $t=1,2,\cdots,8$ 的数据序列 x_t 如下：7.0，6.8，7.3，7.5，8.3，3.8，3.3，2.4．(1)求均值 $\bar{x}$，(2)求 $w_t=x_t-\bar{x}$，(3)用 AR(1)模型拟合 w_t(即 $w_t=\varphi_1 w_{t-1}+\varepsilon_t$)，求 φ_1 的估计值；(4)计算 $\varepsilon_t(t=1,2,\cdots,8)$，(5)作 w_t 对 w_{t-1} 的图形，并在同一图上画出直线 $\hat{w}_t=\varphi_1 w_{t-1}$，(6)计算 σ_x^2，(7)作 ε_t 对 ε_{t-1} 的图形，(8)作 ε_t 对 w_{t-1} 的图形，(9)计算 ε_t 和 ε_{t-1} 的相关函数，(10)计算 ε_t 和 w_{t-2} 的图形．

2. 一个古典的时间序列模型，即 Beveridge 小麦价格指数序列，它是滑动平均模型最早的应用之一．Wold(1938)证明了 MA(2)模型对这 100 个观察序列是运用的，该模型可写成

$$x_t=\varepsilon_t+0.9\varepsilon_{t-1}+0.3\varepsilon_{t-2}$$

如果应用 ARMA(2,1)对序列进行拟合，可得

$$x_t=0.55x_{t-1}-0.21x_{t-2}+\varepsilon_t+0.36\varepsilon_{t-1}$$

两个模型中有一个错了吗？如果有，是哪一个错了？如果两个都对，为什么？说明你的理由．如果有人认为一个纯 AR 模型运用于这组数据，你能说这模型的表达式应该是什么？

3. 时间序列如下：

289	289	289	286	288	287	288	292	291	291
292	296	297	301	304	304	303	307	299	296
293	301	293	301	295	284	286	286	287	284
282	278	281	278	277	279	278	270	268	272
273	279	279	280	275	271	277	278	279	283

计算编相关函数 a_{11},a_{22},a_{33}，并检验假设 $a_{kk}=0,k>1$．

4. 导出 MA(1)模型 $x_t=\varepsilon_t+\theta\varepsilon_{t-1}(-1<\theta<1)$ 的偏相关表达式 a_{kk}，并说

明它是衰减的．

5．设由三个 x_t 序列观察值中已分别获得以下结果

(1) $\bar{x}=0.03, \sigma_x^2=3.34, N=200$

	1	2	3	4	5
ρ_k	−0.800	0.670	−0.518	0.390	−0.310
a_{kk}	−0.800	0.085	0.112	−0.046	−0.061

(2) $\bar{x}=-0.34, \sigma_x^2=1.34, N=200$

	1	2	3	4	5
ρ_k	0.449	−0.056	−0.023	0.028	0.013
a_{kk}	0.449	−0.324	0.218	−0.118	0.077

(3) $\bar{x}=-0.05, \sigma_x^2=2.32, N=200$

	1	2	3	4	5
ρ_k	−0.719	0.337	−0.083	0.075	−0.088
a_{kk}	−0.719	−0.375	−0.048	0.239	0.173

分别画出 ρ_k 和 a_{kk}的图形，并对各自的模型作出估计．

6．有一组数据用 AR(5)模型拟合得

$$(1+0.6B+0.06B^2-0.066B^3+0.0726B^4-0.08B^5)x_t=\varepsilon_t$$

令欲用 ARM2(2,1)模型作最小二乘拟合，求 a_1, a_2, b 的初始估计值(提示：采用逆函数法)．

7．某条河流的一个水文站从 1915 年到 1973 年记录的每年最大径流量如下(共 59 个数据)．

t	1	2	3	4	5	6	7	8	9	10
x_t	15600	89600	10400	10600	10800	9880	9850	10900	8810	9960
t	11	12	13	14	15	16	17	18	19	20
x_t	12200	7510	8640	6380	6810	8820	14400	7440	7240	6830

t	21	22	23	24	25	26	27	28	29	30
x_t	11000	7340	9260	5290	9130	7480	6980	9650	7260	8750
t	31	32	33	34	35	36	37	38	39	40
x_t	9900	9310	9040	7310	8850	7840	10700	6190	9610	7580
t	41	42	43	44	45	46	47	48	49	50
x_t	9990	6150	8250	6030	8080	6180	9630	9490	2340	11100
t	51	52	53	54	55	56	57	58	59	
x_t	5090	10900	6490	12600	6640	7430	6760	10000	9300	

(1)求样本自样关函数;(2)求偏相关函数;(3)确定模型类型、阶数、估计参数;(4)利用1971年与1972年径流量计算1973年径流量的预报值,并与真实值比较,计算预报误差 $e_k=x_{k+1}-\hat{x}_{k+1}$.

8. 试对下述测量值建立数学模型,并用图示方法比较数据曲线和预报结果(一步预报).

t	1	2	3	4	5	6	7	8	9	10
x_t	1.881	2.779	3.792	4.160	3.644	4.103	4.034	2.219	1.265	0.5087
t	11	12	13	14	15	16	17	18	19	20
x_t	0.3356		−2.162		−1.106		−1.456		1.203	
		−1.594		−0.8087		−1.560		−0.193		1.615
t	21	22	23	24	25	26	27	28	29	30
x_t	2.654		1.153		0.7413		0.9099		0.3442	
		1.724		0.2996		1.723		1.158		−0.6364
t	31	32	33	34	35	36	37	38	39	40
x_t	−0.2362		0.6017		3.202		4.946		2.314	
		−0.0850		1.620		4.657		3.339		1.989

t	41	42	43	44	45	46	47	48	49	50
x_t	1.924	1.579	1.125	1.727	3.186	3.604	2.576	−0.7390	−1.408	−0.7285

t	51	52	53	54	55	56	57
x_t	0.3135	1.286	2.687	4.350	5.429		

9. 利用§7.3中叙述的方法，对例8-2局部脑血流量问题进行讨论．

第三部分　仿真与其他方法

第九章　计算机仿真

仿真(也称为模拟)就是在计算机上模仿各种实际系统的运行过程,在整个运行时间内,对系统状态的变化进行观察和统计,从而得到对系统基本性能的估计或认识. 仿真方法实质上是统计估计方法,等效于抽样试验. 对更接近于真实的复杂系统,常常用仿真技术来研究其行为,而这些系统一般不大可能得到解析解.

仿真通常用于以下两种情况:

1°当系统中存在众多随机因素,难以构造经典的数学模型和用解析法求解时,可以利用系统仿真面向问题和面向过程的特点,建立仿真模型,并通过仿真运行得到系统的动态特性.

2°对于多数复杂系统、贵重系统或未来系统等,由于运行费用过高或无法作实际运行,也可借助系统仿真,在没有实际过程介入的情况下,通过仿真模型对系统的行为进行仿真,以得到评价系统所需要的各种参数.

为了模拟实际系统的行为过程,必须构造出能够反映实际系统中基本要素和各要素之间规律的本质关系的仿真模型. 模型中的状态变量应能表示系统的基本特性. 例如排队系统中最基本的要素是顾客和服务机构,它们是由有一定概率分布的到达过程和服务过程联系起来的,并且可以用系统中和队列中的平均顾客数、顾客在系统中和队列中的平均等待时间来表示系统的性能,从而可以构造成最典型的离散、动态、随机的仿真模型.

根据系统的特点,仿真主要分为离散系统仿真和连续系统仿

真．离散系统仿真中常以一个或一组状态变量来表示系统的状态，随着时间的推进和随机事件的出现，这些变量离散地发生着变化. 因此离散系统的仿真模型通常表现为变量在不同变化阶段的行为规则. 例如，某生产线上，零件按泊松分布到达，工序Ⅰ和工序Ⅱ的加工时间分别为正态分布和 β 分布的随机变量，现在要了解零件入库的分布特征. 在这个面向过程的仿真中，零件是流动实体，以一定概率分布输入系统，形成流动实体流，而当实体流经过系统内部各个环节（固定实体），如工序Ⅰ和工序Ⅱ时，将产生一系列事件，记录下事件发生的时刻、事件发生后状态变量的变化情况，并对流动实体的数量、延迟时间以及系统性能参数进行统计. 在仿真过程结束时能够提供系统的基本参数（如服务强度、零件平均到达时间等）的数学期望、方差、最大值、最小值和概率分布曲线等.

对连续系统仿真，首先要建立描述连续系统的数学模型. 一个连续系统的模型通常用微分方程、状态方程、传递函数以及系统的结构图等形式表示. 为了将上述数学模型进行数字仿真，需要将它化成便于在计算机上执行运算的离散形式的模型，如差分方程、离散时间状态方程、离散框图等. 通过数值计算，可将整个系统的动态特性以及系统中一些参数变化对系统的动态特性的影响，计算出来. 例如在第六章里河流中污染物扩散的问题，就是分布参数系统的仿真的例子.

这里着重讨论离散系统仿真. 随机型离散系统仿真的关键在于产生所需要的伪随机数.

§9.1 伪随机数发生器

随机数是从一定概率分布总体中进行随机抽样时随机变量所取的数值. 随机数发生器则是在计算机上产生规定分布随机数流的方法或过程.

要产生一定分布的随机数，通常先要产生均匀分布的随机数，然后才能从所需要分布的概率密度函数或累积分布函数中产生出相应的随机数. 因此，均匀分布的随机数是产生其他分布随机数的基础.

一、伪随机数发生器

在随机数发生器中，经常用到的是标准均匀分布的随机数，记为$U(0,1)$. 从理论上说，利用算法过程产生的“随机数”，并不具有真正的随机性，因此，人们将这类由算法过程产生的随机数称为“伪随机数”. 然而，如果仔细地设计发生器的算法和合理地选择参数，是可以产生出一系列很接近$U(0,1)$分布的随机数，满足工程和管理上的实际要求的.

产生伪随机数的方法很多. 下面仅介绍三种.

1. 中值平方法

首先选择一个P位整数作为“种子数”或称初值. 将种子值平方后，取其中间P位数值作为下一个种子值，并对此数进行规格化处理，使之成为P位有效数字且小于1的实数值，作为产生出来的第一个伪随机数. 依此类推可以得到一系列随机数，形成随机数流.

例如，取$P=2$，$x_0=76$(种子数)

则 $x_0^2=76^2=5776$， $x_1=77$， $u_1=0.77$

$x_1^2=77^2=5929$， $x_2=92$， $u_2=0.92$

…

2. 中值乘积法

先取任意两个p位整型奇数，其中一个作为种子数，另一个作为乘数. 将种子数与乘数相乘，得到小于或等于$2p$位的整奇数. 取其中间p位数进行规格化处理作为所产生的伪随机数，再取其最右端p位数作为下一乘数，再与种子数相乘可以得到第二个$2p$位整奇数. 依此类推可得到相应的随机数流，它可以较好地

产生符合均匀分布的随机数.

例如，取 $p=4$，种子数=5167，第一个乘数=3729

表 9-1

乘数值	种子数与乘数之积	产生的随机数
3729	19267743	0.2677
7743	40008081	0.0080
…	…	…

3. 线性同余法

目前在离散系统仿真中应用最广泛的是线性同余法，其算法过程如下.

令 Z_0 为种子数，Z_i 为第 i 个中间值，a 为常数，c 为增量(常数)，m 为模(取充分大的整数值).

设 $Z_i=(aZ_{i-1}+c)\mathrm{mod}\,m$，其含意是将 $(aZ_{i-1}+c)$ 除以 m 并取其余数作为 Z_i，或

$$Z_i=(aZ_{i-1}+c)-\left\lfloor\frac{(aZ_{i-1}+c)}{m}\right\rfloor\cdot m,$$

其中$\lfloor\quad\rfloor$表示取较小的整数，显然有

$$0\leqslant Z_i\leqslant m-1$$

令 $u_i=\dfrac{Z_i}{m}$，则 u_i 在 $(0\sim\dfrac{m-1}{m})$ 之间变化. 当 m 充分大时，则 u_i 可认为在(0,1)之间取值.

例如，取 $m=16, a=5, c=3, Z_0=7$

即 $$Z_i=(5Z_{i-1}+3)\ \mathrm{mod}\ 16$$

则有

$i=0$	$Z_0=7$	
$i=1$	$Z_1=6$	$u_1=0.375$
$i=2$	$Z_2=1$	$u_2=0.063$
$i=3$	$Z_3=8$	$u_3=0.500$
…	…	…

二、产生规定分布的随机变量

仿真过程中需要用到各种不同类型的概率分布，因而需要有各种概率分布的随机数发生器．一般仿真语言中均提供了负指数分布、均匀分布、正态分布、对数正态分布、爱尔朗分布、β分布、γ分布、三角分布、韦伯尔分布、二项分布、泊松分布等随机数发生器，供用户调用．

对于任意分布的随机变量X，令$F(x)$为其累积分布函数．设$y=F(x)$，则Y也是一个随机变量，且$G(y)$为Y的累积分布函数，Y在$[0,1]$内取值，则

$$G(y)=P\{Y\leqslant y\}=P\{F(x)\leqslant y\}$$
$$=P\{X\leqslant F^{-1}(y)\}=y$$

因此，随机变量Y的概率密度函数$g(y)$为

$$g(y)=\frac{\mathrm{d}G(y)}{\mathrm{d}y}=\frac{\mathrm{d}y}{\mathrm{d}y}=1$$

按定义有

$$0\leqslant y=F(x)\leqslant 1$$

即Y是$(0,1)$区间内的均匀分布随机变量．

以上分析表明，若$F(x)$是任意分布的随机变量X的累积分布函数，则$y=F(x)$所对应的随机变量Y是$(0,1)$区间内均匀分布的随机变量，并与X的分布无关．

以上结论是用逆变法产生规定分布随机变量的依据．

1．逆变法

逆变法的特点是利用任意随机变量的累积分布函数服从均匀分布这一性质，先由伪随机数发生器产生一组独立的$U(0,1)$随机数u_i，令$F(x_i)=u_i$，则每一个u_i对应于一个x_i，即$x_i=F^{-1}(u_i)$，故x_i就是$f(x)$的一个随机数．

例如负指数分布的概率密度函数为

$$f(x)=\lambda e^{-\lambda x}\quad (x\geqslant 0)$$

则
$$F(x)=\int_0^x \lambda e^{-\lambda t}dt=1-e^{-\lambda x}$$
设 u_i 为 $U(0,1)$随机数，使 $u_i=F(x_i)=1-e^{-\lambda x}$，由于 u_i 和$(1-u_i)$具有相同的 $U(0,1)$分布特性，可以令 $u_i=e^{-\lambda x_i}$，或 $\ln u_i=-\lambda x_i$，$x_i=-\frac{1}{\lambda}\ln u_i$，则 x_i 就是所求负指数分布的随机数.

由于逆变法必须将所求随机变量的累积分布函数写成逆函数形式，这对于多数概率分布如正态分布、r 分布等仍是十分困难的，因此还需要其他可行的方法.

2. 卷积法

有些常用的概率分布，其随机变量可以表示为若干其他分布随机变量之和，而这些随机变量都是相互独立的、同分布的，并能较方便地产生随机数，这时可采用卷积法. 设 X 为某一分布的随机变量，$Y_1,Y_2,\cdots,Y_k$ 为 k 个独立、同分布的随机变量，若 $Y_1+Y_2+\cdots+Y_k$ 所构成的随机变量的分布就是随机变量 X 的分布，则 X 的分布称为 Y_i 的 k 次卷积.

设 $P\{X\leqslant x\}=P\{(Y_1+Y_2+\cdots+Y_k)\leqslant x\}=F(x)$为 X 的累积分布函数，$G(y)$为 Y 的累积分布函数，则卷积法的算法过程为：

1° 用累积分布函数 $G(y)$分别对 $Y_1,Y_2,\cdots,Y_k$ 各产生一个随机数.

2° $X=Y_1+Y_2+\cdots+Y_k$，就是所求概率分布的随机数.

例如爱尔朗分布的概率密度函数为
$$f(x)=\frac{1}{(k-1)!\ \beta^k}e^{-\left(\frac{x}{\beta}\right)}x^{k-1},k>1 \text{ 为整数}.$$
这相当于 k 个参数为 $k\beta$ 的负指数分布随机变量之和构成的概率分布.

令 Y_j 为参数为 $k\beta$ 的负指数分布随机变量，先用逆变法求得 $Y_j=-\frac{1}{k\beta}\ln u_j$，$j=1,2,\cdots,k$，然后令 $X=\sum_{j=1}^{k}\left(-\frac{1}{k\beta}\ln u_j\right)=-\frac{1}{k\beta}\ln\left(\prod_{j=1}^{k}u_j\right)$，其中 u_j 为 $U(0,1)$ 分布的随机数，则 X 就是爱

尔朗分布的随机变量.

3. 取舍法

以上所述各种产生规定概率分布随机数的方法都与分布特征直接有关，因此称为直接法. 取舍法则不然，它不是直接从累积分布函数或随机变量的组成中产生随机数，而是蒙特卡罗仿真的一种随机抽样方法.

设 $f(x)$ 为要求产生随机数的概率密度函数，选取 $t(x) \geqslant f(x)$，$x \in [a,b]$，

因
$$C = \int_a^b t(x)\mathrm{d}x > \int_a^b f(x)\mathrm{d}x = 1$$

定义
$$r(x) = \frac{t(x)}{C}$$

则
$$\int_a^b r(x)\mathrm{d}x = \frac{1}{C}\int_a^b t(x)\mathrm{d}x = 1$$

即 $r(x)$ 为概率密度函数.

取舍法产生随机数的算法过程为

1°从密度函数 $r(x)$ 中产生一个随机数 y

2°从 $U(0,1)$ 分布中产生随机数 u

3°判断：若 $u \leqslant \dfrac{f(y)}{t(y)}$，则"取"，置 $x = y$，返回.

若 $u > \dfrac{f(y)}{t(y)}$，则"舍"，返回 1°.

按此算法即可得到规定分布 $f(x)$ 的随机数.

例如，在(0,1)区间内的 β 分布，其概率密度函数为

$$f(x) = \begin{cases} 60x^3(1-x)^2, & 0 \leqslant x \leqslant 1 \\ 0, & \text{其他} \end{cases}$$

求得 $f(x)$ 的最大值：$\max f(x) = f(0,6) = 2.0736$.

定义
$$t(x) = \begin{cases} 2.0736, & 0 \leqslant x \leqslant 1 \\ 0, & \text{其他} \end{cases}$$

则
$$r(x) = \frac{t(x)}{C} = \frac{2.0736}{\int_0^1 t(x)\mathrm{d}x} = 1$$

因此，$r(x)$是[0,1]的均匀分布．按取舍法，可得 β 分布的随机数．

§9.2 仿真输出数据的分析

对仿真的输出数据进行统计分析是为了估计系统的性能或比较两个或多个系统方案的性能，在离散系统仿真中，根据不同性质的问题或仿真实验的目的，可将仿真过程分为两类：

终态仿真——每次仿真运行的终止条件为：规定的仿真时间或规定的事件发生．其特点是必须规定初始条件，同时必须定义终止时间或终止事件．

稳态仿真——仿真运行持续足够长的时间，这时系统的性能与仿真的初始条件已无关，并达到系统平稳状态的性能和平稳分布．

对于不同类型的仿真，其输出数据的分析具有不同的特点．

一、性能测度及其估计

设系统的真实参数所对应的随机变量的数学期望为 $E[Y]$，仿真的输出形式为$\{Y_1,\cdots,Y_n\}$或$\{Y(t),0\leqslant t\leqslant T\}$，它们都是系统的随机样本．性能估计就是根据随机样本的数据来估计与真实参数的差别和代表程度．性能估计可分为点估计和区间估计两个方面．

1．点估计

通过样本数据所得到的样本均值与方差来估计真实参数的期望与方差．

设仿真试验的样本为$\{Y_1,Y_2,\cdots,Y_n\}$或$\{Y(t),0\leqslant t\leqslant T\}$，则定义

$$\overline{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i \text{ 或 } \overline{Y}_t=\frac{1}{T}\int_0^T Y(t)\mathrm{d}t$$

若 $E[\overline{Y}]=E[Y]$或$E[\overline{Y}_t]=E[Y_t]$，则称$\overline{Y}$和$\overline{Y}_t$为

$E[Y]$ 或 $E[Y_i]$ 的无偏估计.

定义 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\overline{Y})^2$,则 S^2 是 $V[Y]$的无偏估计.

2. 区间估计

根据样本数据对系统性能的期望 $E[Y]$或方差 $V[Y]$所在区间的估计.

令 $V[\overline{Y}]$为点估计 $\overline{Y}$ 的真实方差,s 为对 $V[Y]$的点估计,则定义

$$t=\frac{\overline{Y}-E[Y]}{\sqrt{\frac{s^2}{n}}}$$

其中 $Y_1,Y_2,\cdots,Y_n$ 是用独立随机数流产生的仿真输出响应,可以认为是独立同分布的随机变量. 可以证明 t 是一个自由度为 $n-1$ 的 t 分布令 $P_n(t)$为其密度函数. 若取 λ 值使得

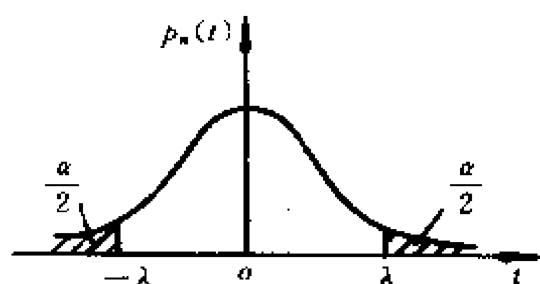


图 9-1 自由度为 $n-1$ 的 t 分布

$$P\{-\lambda\leqslant t\leqslant\lambda\}=\int_{-\lambda}^{\lambda}p_n(t)\mathrm{d}t=1-\alpha$$

则
$$P\{t\geqslant\lambda\}=P\{t\leqslant-\lambda\}=\frac{\alpha}{2}$$

令
$$\lambda=t_{\alpha/2,n-1}$$

则
$$P\left\{\left|\frac{\overline{Y}-E[Y]}{S/\sqrt{n}}\right|\leqslant t_{\alpha/2,n-1}\right\}=1-\alpha$$

即 $|\overline{Y}-E[Y]|\leqslant t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$的概率为 $1-\alpha$,其中 α 称为显著水平,$1-\alpha$ 称为置信度.

因此 $E[Y]$的 $100(1-\alpha)\%$的置信区间为

$$\overline{Y}-t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\leqslant E[Y]\leqslant\overline{Y}+t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$$

式中 $t_{\alpha/2,n-1}$ 可直接由 t 分布函数表查得.

二、终态仿真的输出分析

设在 $(0,T_E)$ 中对系统作终态仿真，n 次观测的结果为 $Y_1,Y_2,\cdots,Y_n$. 由于在一次终态仿真运行中，都是用同一随机数流进行仿真，并且每次观测的结果都是下一次运行的初始条件，从而使观测结果形成自相关的数列. 为了保证仿真输出的独立性，可以采用独立随机流作独立重复的仿真运行. 其中重复的意义是指具有相同的初始条件.

如果令 Y_{ri} 为第 r 次重复运行中第 i 次观测结果，则对于同一个 r，$Y_{r1},Y_{r2},\cdots,Y_{rn}$ 是自相关序列，而对于 Y_{ri} 和 Y_{si}，$r\neq s$，则它们之间是相互独立的. 若定义每次重复运行的均值为 $\overline{Y}_r$，则

$$\overline{Y}_r=\sum_{i=1}^{n}\frac{Y_{ri}}{n}$$

若共进行 k 次重复运行，则 $\overline{Y}_1,\overline{Y}_2,\cdots,\overline{Y}_k$ 是独立、同分布的随机序列. 总的点估计为

$$\overline{Y}=\frac{1}{k}\sum_{r=1}^{k}\overline{Y}_r$$

其方差点估计为

$$V[\overline{Y}]=\frac{S^2}{R}=\frac{1}{R(R-1)}\sum_{r=1}^{k}(\overline{Y}_r-\overline{Y})^2$$

故对于 $E[Y]$ 的 $100(1-\alpha)\%$ 置信区间为

$$\overline{Y}-t_{\alpha/2,R-1}\frac{S}{\sqrt{R}}\leqslant E[Y]\leqslant\overline{Y}+t_{\alpha/2,R-1}\frac{S}{\sqrt{R}}$$

三、稳态仿真的输出分析

为了估计系统达到平稳状态时的性能，可以通过长时间的仿真运行，并产生观测值 $\{Y_1,Y_2,\cdots\}$，但是它们是自相关的随机序

列. 从理论上说，系统的稳态测度 $E[Y]$应为

$$E[Y]=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}Y_i$$

即系统的全部仿真运行，其样本均值应收敛于 $E[Y]$. 但是从经济上考虑进行长时间的仿真是花费很大的. 因此仍应采用重复仿真运行的方法.

在作重复仿真运行时，如果运行长度不够，初始条件可能在数据分析中造成较大的偏差. 为了得到稳态的仿真响应，可从两个方面来消除初始偏差. 第一，从实际模型的工作状态出发，将仿真的初始条件尽可能置于典型的系统状态，以消除或减弱初始偏差. 第二，将仿真运行分成初始阶段和数据采集阶段，如图 9-2 所示

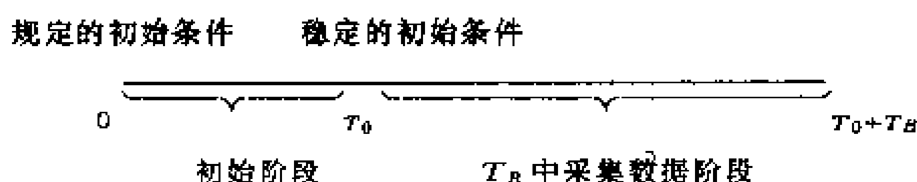


图 9-2 稳态仿真的时间分段

系统在 T_0 时的状态具有一定的稳态代表性，而 T_B 的时间应足够长，以便能充分精确地估计稳态性能.

对于第 r 次重复运行，其均值为

$$\overline{Y}_r(n,d)=\frac{1}{n-d}\sum_{i=d+1}^{n}Y_{ri}$$

即每次运行都删去初始 d 个仿真数据，以消除初始条件的影响，同时，由于每次重复运行均用不同的随机数流，并在 $T=0$ 时置相同的初始条件，则 $\overline{Y}_1(n,d),\overline{Y}_2(n,d),\cdots,\overline{Y}_R(n,d)$是独立、同分布的随机样本.

总的点估计为

$$\overline{Y}(n,d)=\frac{1}{R}\sum_{r=1}^{R}\overline{Y}_r(n,d)$$

若 d 和 n 都选得足够大，使 $E[\overline{Y}(n,d)]=E[Y]$，则 $\overline{Y}(n,d)$就

是 $E[Y]$的近似无偏点估计.

类似地可以得到 $E[Y]$的 $100(1-\alpha)\%$置信区间,但其中

$$s^2=\frac{1}{k-1}\sum_{r=1}^{k}[\overline{Y}_r(n,d)-\overline{Y}(n,d)]^2$$

四、取得规定精度的置信区间

根据一定重复仿真运行次数 R,可以得到 $100(1-\alpha)\%$的置信区间,若要求置信度一定,则置信区间往往不能达到所要求的精度. 如果重复仿真次数过少,则置信区间将变宽,如果重复仿真次数过多,则将造成不必要的计算费用增加. 因此,取得规定精度的置信区间是十分必要的.

定义置信区间的半长为该置信区间的绝对精度,而置信区间与点估计 $\overline{Y}$ 的比值为置信区间的相对精度.

设规定的绝对精度为 ε,于是要求点估计 $\overline{Y}$ 与 $E[Y]$之间的差别不超过规定精度的概率至少为 $1-\alpha$,即

$$P\{|\overline{Y}-E[Y]|<\varepsilon\}\geqslant 1-\alpha$$

若相对精度为 r,则

$$P\left\{\left|\frac{\overline{Y}-E[Y]}{\overline{Y}}\right|<r\right\}\geqslant 1-\alpha$$

设先对系统作 $R_0\geqslant 2$ 次重复仿真运行,则 R_0 次仿真运行的观测结果可得到对总体方差的点估计 S_0^2,由计算置信区间的公式知,应使初始仿真 R_0 次运行的置信区间半长小于或等于规定的精度,即

$$t_{\alpha/2,R-1}\frac{S_0}{\sqrt{R}}\leqslant\varepsilon$$

由上列不等式可解出满足精度要求的最少重复运行次数

$$R\geqslant(t_{\alpha/2,R-1}S_0/\varepsilon)^2$$

显然 $R>R_0$,只要对系统再作 $R-R_0$ 次重复仿真运行,所得到的置信区间半长必将满足规定精度 ε 的要求. 以上过程用相对精度

计算也可得到相同的结果.

但是,上述结果是假设样本方差的点估计 S_0 不随 R 的增大而变化得到的,实际上样本方差的点估计随 R 的增大而减小,因此还需作精度的修正.

另一种保证精度的方法是在初始重复仿真次数 R_0 的基础上,每增加一次仿真运行就进行一次精度验算,直到满足精度要求时仿真结束. 这种序贯迭代方法可嵌入离散系统仿真软件中,作为停止仿真的条件.

仿真既是利用模型来模仿真实系统,当然两者不可能完全等同,因此在仿真工作中需要确定所建模型是否能有效地模仿真实系统. 人们在仿真中并不说模型绝对有效或无效,而是说模型与相应真实系统一致程度(称为仿真信度)的高或低.

确定仿真信度的常用方法是分别根据真实系统的观测数据和相应模型的输出数据,计算一个或几个相应统计量,然后用所得到的模型统计量与真实系统的统计量对比,为某些仿真研究提供模型是否适用的有价值的信息. 另一种方法是利用数理统计方法对来自系统的 m 个独立数据集合和来自模型的 n 个独立数据集合,进行假设检验和置信区间的构造,这一方法更为可靠.

应用仿真方法的主要缺点是它只能得到系统的特解,而不是通解. 每运行一次,只能得到一个在特殊设定条件下的解. 为了找到所有条件下的解答,就得在不同条件下,重复进行仿真,特别是需要求得更一问题的极值时,需要经过多次仿真运算,而且很难确定是局部还是总体极值. 尽管仿真方法有其缺点,但由于应用解析方法求解问题的范围毕竟有限,而仿真方法通过建立系统的合理的模型,配合相应有效的计算手段,就能以一种经济的方式进行各种试验,在某些特性上"再现"真实系统,所以越来越受到人们的重视.

§9.3 实　　例

下面两个例子取自于美国大学生数学模型比赛的试题. 例 9-1 是 1989 年 B 题,解答取材于 Harvey Mudd 学院队的论文,例 9-2 是 1992 年 B 题,解答取材于 Washington University 队的论文.

例 9-1 机场通常都是用"先来后到"的原则来分配飞机跑道,即当飞机准备好离开登机口时,驾驶员电告地面控制中心,加入等候跑道的队伍.

假设控制塔可以从快速联机数据库中得到每架飞机的如下信息:

1° 预定离开登机口的时间;

2° 实际离开登机口的时间;

3° 机上乘客人数;

4° 预定在下一站转机的人数和转机的时间;

5° 到达下一站的预定时间.

又设共有 7 种飞机,载客量从 100 人起以 50 人递增,载客最多的一种是 400 人.

试开发和分析一种能使乘客和航空公司双方满意的数学模型.

问题分析:问题考虑到机场和乘客的满意,是一个多目标优化问题,(在折换成费用时,可转为)可加性条件下的单目标. 飞机是否起飞可归结为 0-1 规划问题.

1. 模型假设

1° 机场仅有一条跑道供飞机起飞,任何飞机起飞占用跑道的时间都相同. 设这个时间为 Δ,于是时间被离散化为间隔为 Δ 的窗口.

2° 第 i 架飞机在第 j 窗口起飞的费用与已经起飞的飞机无关. 这一假定使得给定的一串飞机起飞的总费用是线性函数.

3° 对于每架飞机，存在它可以延迟起飞的最晚时间 τ. 若起飞时间不迟于 τ，飞机加速飞行仍可按时到达下一站；若起飞时间迟于 τ，则飞机要以最高速度飞行，而且即使这样，所有要在下一站转机的乘客也要误机(无法转乘预订的飞机).

4° 所有要转机乘客的误机损失费相同.

2. 模型的分析与设计

假定 $t=0$ 时有 n 架飞机请求起飞，机场控制塔要设计一个起飞次序，即为每架飞机安排一个窗口，使得按照这个次序起飞时总费用最小. 总费用包括两部分，一是当飞机比预定时间延迟起飞时航空公司需付的附加费，二是飞机迟飞引起乘客不满意而折合的损失费.

记 c_{ij} 是第 i 架飞机被安排在第 j 窗口起飞时，这架飞机承担的上述两部分费用. 定义

$$x_{ij}=\begin{cases}1 & \text{若第 } i \text{ 架飞机安排在第 } j \text{ 窗口起飞}\\ 0 & \text{否则}\end{cases}$$

显然，对于任一个起飞次序，总费用为

$$c=\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}x_{ij} \tag{1}$$

为保证对于每架飞机有且仅有一个窗口，约束条件是

$$\sum_{j=1}^{n}x_{ij}=1, i=1,2,\cdots,n; \tag{2}$$

$$\sum_{i=1}^{n}x_{ij}=1, j=1,2,\cdots,n. \tag{3}$$

据假定 c_{ij} 与 x_{ij} 无关，c 是 x_{ij} 的线性函数. 于是寻求最优起飞次序归结为：确定 $x_{ij}, i,j=1,2,\cdots,n$，使在条件(2)、(3)下目标函数(1)达到最小.

这是一个 0-1 规划问题，在 n 不太大的情况下现成的算法和软件包可供实际应用. 这样，建模的关键在于确定 c_{ij}.

对于第 i 架飞机，记预定起飞时间为 t_1^i，$t_1^i>0$ 表示请求起飞时

间($t^i=0$)早于预定起飞时间,$t_1^i<0$ 则表示请求起飞时间晚于预定时间.若安排它在第 j 窗口起飞,这个起飞时间记作 t^i,显然 $t^i=(j-1)\Delta$. 按照假设(3),记可以延迟起飞的最晚时间为 τ^i,它预定到达下一站的时间为 t_2^i,飞行距离为 d^i,飞行的正常速度和最大速度分别为 v^i 和 v_m^i,根据 τ 的意义应有

$$d^i=(t_2^i-t_1^i)v^i=(t_2^i-\tau^i)v_m^i$$

于是 τ^i 可由已知数据 t_2^i,t_1^i,v^i,v_m^i 确定:

$$\tau^i=t_2^i-\frac{(t_2^i-t_1^i)v^i}{v_m^i}$$

记第 i 架飞机上的乘客数为 p^i,在下一站转机的乘客数为 q^i. c_{ij}由以下情况决定.

当 $t<t_1^i$ 时,$c_{ij}=\infty$. 表示不允许在预定起飞时间之前起飞.

当 $t_1^i\leqslant t\leqslant\tau^i$ 时,飞机加速飞行引起燃料消耗,航空公司所付的附加费记为 $f_1(t)$,为简单起见设它与飞行距离 d^i 和延迟的时间 $t-t_1^i$ 成正比,考虑到 d^i 与 $t_2^i-\tau^i$ 成正比,所以有

$$f_1^i(t)=k(t_2^i-\tau^i)(t-t_1^i)$$

其中 k 为比例系数.乘客的不满意程度将随着延迟的时间增加而迅速增长,设每个乘客的不满意程度相同,记这种不满意折合的损失费为 $g_1^i(t)$,设

$$g_1^i(t)=a^i[e^{\alpha^i(t-t_1^i)}-1]p^i$$

其中,a^i,α^i 为比例系数,则

$$c_{ij}=f_1^i(t)+g_1^i(t).$$

当 $t>\tau^i$ 时,燃料消耗引起的附加费为 $f_1^i(\tau)$(以最大速度 v_m 飞行,燃料消耗不再随 t 增加),所有乘客不满意折合的损失费为 $g_1^i(t)$,除此之外还需考虑误机引起的费用.设航空公司赔偿误机乘客的费用为

$$f_2^i(t)=r^iq^i$$

r^i 为比例系数;误机乘客的抱怨折合的损失费设为

$$g_2^i(t)=b^i q^i$$

b^i 是比例系数. c_{ij}是 $f_1^i(t)$,$g_1^i(t)$、$f_2^i(\tau)$,$g_2^i(t)$之和.

综上所述,

$$c_{ij}=\begin{cases}\infty,0\leqslant t<t_1^i\\ f_1^i(t)+g_1^i(t),t_1^i\leqslant t\leqslant \tau^i\\ f_1^i(\tau)+g_1^i(t)+f_2^i(t)+g_2^i(t),t>\tau^i\end{cases}$$

在 c_{ij}的表达式中,t_1^i,t_2^i,p^i,q^i 由数据信息给出,Δ,v^i,v_m^i,k^i,r^i 是可以预先确定的常数,a^i,b^i,α^i 是自由参数,这些将乘客主观的不满意程度折合成损失费的比例系数难以精确估量,可根据经验估计.

3. 模型检验

建立的模型是否能够应用,在很大程度上取决了当自由参数作微小改变时,最优解是否变化很大,以及各个自由参数的改变对结果的影响程度.

为了分析最优解对参数的敏感性,首先考虑相应的线性规划问题.

记(1)式中的 c 为$-z$,构造线性规划问题:

$$\max\ Z=\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}x_{ij}$$

$$\text{s.t.}\ \sum_{j=1}^{n}x_{ij}=1,i=1,2,\cdots,n$$

$$\sum_{i=1}^{n}x_{ij}=1,j=1,2,\cdots,n$$

$$x_{ij}\geqslant 0$$

若(1)式有可行解,则一定存在决策变量 x_{ij}取整数值的最优解. 注意到前两个约束条件,x_{ij}只能取值 0 和 1,于是整数规划问题等价于原来的 0-1 规划. 将问题(1)简记为

$$\max\ Z=\boldsymbol{C}^T\boldsymbol{X}$$

$$\text{s.t.}\ \boldsymbol{AX}=\boldsymbol{b},\boldsymbol{X}>\boldsymbol{0}. \tag{4}$$

设问题(4)的最优解为 X^*,研究 C 有微小改变 Δc 时的影响.

问题(4)的对偶问题是

$$\min W = b^T Y$$
$$\text{s.t.}\ \ YA \geqslant C$$

设其最优解为 Y^*,因为当 C 变为 $C^1 = C + \Delta C$ 时,X^* 作为(4)式的可行解,根据线性规划的对偶理论,当且仅当 $Y^* A \geqslant C^2$ 时,即 $\Delta C \leqslant Y^* A - C$ 时,X^* 还是(4)式的最优解.

进一步说,即使 x^* 不再是最优解,新的目标函数为

$$Z^1 = C^{1T} X^* = Z + (\Delta C)^T X^*,$$

即 Z 的改变量是 ΔC 的线性函数.

综上所述,当费用 c_{ij}发生微小变化时对最优解和目标函数值的影响是微小的.

c_{ij}的不确切性主要源于自由参数 a,b,α,设它们的估计值和变化范围是

$$\hat{a} \pm \sigma_a, \quad \hat{b} \pm \sigma_b, \quad \hat{\alpha} \pm \sigma_\alpha.$$

则 c_{ij}的变化范围 $\sigma_{c_{ij}}$可由下式计算:

$$\sigma_{c_{ij}}^2 = \sigma_a^2 \left(\frac{\partial c_{ij}}{\partial a} \right)^2 + \sigma_b^2 \left(\frac{\partial c_{ij}}{\partial b} \right)^2 + \sigma_\alpha^2 \left(\frac{\partial c_{ij}}{\partial \alpha} \right)^2$$

而上面的偏导数容易从 c_{ij}的表达式算出.

最后,不难知道目标函数值的变化范围 σ_z 由

$$\sigma_z^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{c_{ij}}^2 x_{ij}^2$$

决定.

4. 计算机模拟

理论模型的约束比仿真模型少得多.为简洁起见,增加以下假设:

1° 最多有三架飞机准备起飞,若只有两架飞机准备起飞,则插入一架虚拟飞机,其耗费系数全为零.

2° 直观指定模型参数,实际上这些参数值可由经验或调查确

定.

3° 每一起飞窗口为一分钟长；没有飞机降落在跑道上；任一乘客的转机费用为 350 美元；未起飞的预订飞机的乘客的损失是误点 15 分钟乘客的两倍.

仿真例 1：考虑一种简单情况，在早上 6：00，有三架飞机 A、B、C 准备起飞，飞往三个城市，并都将在早上 7：20 到达目的地，仿真结果如表 9-2：

表 9-2　三架飞机仿真结果(1)

飞机	乘客数/转机人数	耗费矩阵			解		
A	350/100	0.00	0.48	0.97	0	1	0
B	100/100	0.00	0.41	0.83	0	0	1
C	400/100	0.00	0.50	1.00	1	0	0

最优起飞顺序为 C，A，B，即在所有情况一致时，乘客最多的飞机排为第一.

仿真例 2　考虑在飞机 C 起飞时，飞机 D 要求起飞，D 已误点 10 分钟，注意到此时 A、B 均误点 1 分钟，仿真结果如表 9-3 所示.

表 9-3　三架飞机仿真结果(2)

飞机	乘客数/转机人数	误时	耗费矩阵			解		
D	210/140	10 分钟	0.82	0.91	1.00	1	0	0
A	100/100	1 分钟	0.07	0.15	0.22	0	0	1
B	350/100	1 分钟	0.09	0.17	0.26	0	1	0

最优排序为 D，A，B，即误点最长的飞机排序第一.

仿真例 3　两分钟过去了，飞机 D、A 都已起飞，此时飞机 B 误点 3 分钟，另一架飞机 E 也已准备起飞. 设 E 已在时间表上，而晚点的耗费为 450 美元/分钟. 仿真结果如表 9-4 所示.

表 9-4 三架飞机仿真结果(3)

飞机	乘客数/转机人数	误时	耗费矩阵			解		
B	100/100	3 分钟	0.60	0.80	1.00	0	1	0
E	122/89	0	0.00	0.28	0.56	1	0	0
X	0/0	0	0.00	0.00	0.00	0	0	1

看起来，似乎 B 应在 E 前起飞，但由于 E 高速飞行的耗费和乘客人数，更好的方案是 E 在 B 前起飞.

例 9-2 为沿海地区服务的电力公司必须具备应急系统来处理风暴引起的电力中断. 这样的系统需要由估计修复的时间、费用和由客观准则判定的停电的"价值"构成的数据输入，过去 HECO 电力公司曾因缺乏优先方案而遭受传播媒介的批评.

假设你是 HECO 电力公司顾问，HECO 具有一个实时处理，通常包含下述信息的服务电话的计算机数据库：

报修时间，需求者类型，估计的受害人数以及停电地点 (x, y).

工程队调度所位于 (0,0) 和 (40,40)，其中 x, y 以公里为单位. HECO 的服务区域在 $-65<x<65$ 和 $-50<y<50$ 之内. 因为有极好的道路网络，该地区完全都市化了. 工程队只是在上班和下班时必须回调度所. 公司的政策是假若停电的设施是铁路或医院，只要有工程队可派就立即处理，其他情形都要等暴风雨离开这一地区后才开始工作.

HECO 雇你为表 9-5 所列的暴风雨修复请求和表 9-5 所列的维修能力建立客观准则和安排工作计划. 注意，第一个电话是 4：20(早上)接到的，暴风雨是早上 6：00 离开该地区，还要注意很多停电是当日很迟才报修的.

HECO 出自自身的目的需要一份技术报告和一份用外行术语写就的"执行摘要"，可提交新闻媒介. 他们希望有对将来的建议. 为决定你的优先计划安排系统，你还需作一些附加的假设，详

述这些假设.将来你可能希望有附加的数据,如果是,详述这些需要的信息.

表 9-5 风暴修复请求

时间(a.m)	位置	类型	受影响人数	估计修复时间(一队所需小时数)
4:20	(-10,30)	事业(有线电视)	?	6
5:30	(3,3)	住宅	20	7
5:35	(20,5)	事业(医院)	240	8
5:55	(-10,5)	事业(铁路系统)	25 名工人 75000 乘客	5
6:00	风暴离开	工程队可以派出		
6:05	(3,30)	住宅	45	2
6:06	(5,20)	区域*	2000	7
6:08	(60,45)	住宅	?	9
6:09	(1,10)	政府(市政厅)	?	7
6:15	(5,20)	事业(购物中心)	200 工人	5
6:20	(5,-25)	政府部门(消防)	15 工人	3
6:20	(12,18)	住宅	350	6
6:22	(7,10)	区域*	400	12
6:25	(-1,19)	工业(报业公司)	190	10
6:40	(-20,-19)	工业(工厂)	395	7
6:55	(-1,30)	区域*	?	6
7:00	(-20,30)	政府(高中)	1200 学生	3
7:00	(40,20)	政府(小学)	1700	?
7:00	(7,-20)	事业(饭店)	25	12
7:00	(8,-23)	政府(警察局、监狱)	125	7
7:05	(25,15)	政府(小学)	1900	5

续表

时间(a,m)	位置	类型	受影响人数	估计修复时间(一队所需小时数)
7:10	(−10,−10)	住宅	?	9
7:10	(−1,−2)	政府(学院)	3000	8
7:10	(8,−25)	工业(电脑制造)	450 工人	5
7:10	(18,55)	住宅	350	10
7:20	(7,35)	区域*	400	9
7:45	(20,0)	住宅	800	5
7:50	(−6,30)	事业(医院)	300	5
8:15	(10,40)	事业(几家商店)	50	6
8:20	(15,−25)	政府(交通灯)	?	3
8:35	(−20,−35)	事业(银行)	20	5
8:50	(47,30)	住宅	40	?
9:50	(55,50)	住宅	?	12
10:30	(−18,−35)	住宅	10	10
10:30	(−1,50)	事业(市中心)	150	5
10:35	(−7,−8)	事业(机场)	350 工人	4
10:50	(5,−25)	政府(消防部门)	15	5
11:30	(8,20)	区域*	300	12

* 区域指二个或多个其他类域的组合.

工程队的情况

- 工程队调度所位于(0,0)和(x_0,y_0);
- 工程队由三个熟练工人组成;
- 工程队只是在上下班时间向调度所报告;
- 工程队上班时全部时间用来做它的调度所指派的工作.工程队通常按常规执行任务.在风暴离开该区域之前,他们只能因紧急情况派出;

· 工程队工作 8 小时后换班；

· 每个调度所指挥六个工程队；

· 工程队一天最多加一班，加班领取一倍半工资.

基本假设：从表 9-4 可见，每中心有 3 名熟练工人每天轮流值班(一个工程队)，紧急修理时每个点有 6 名工人可启用(每个点的 6 个工程队分为三班，每班 8 小时). 另外增加假设：

1° 街道为东西—— 南北向，两互间距离公式为：

$$d_{ab}=|x_a-x_b|+|y_a-y_b|$$

道路状况良好，均匀，路途时间只取决于距离．车辆路途运行费用与工资相比可忽略不计.

2° 每个正常班工资为 10 美元/小时，加班费为 15 美元/小时，工人在未完成一项工作时不得移动地点，每个工人可加班 8 小时，但已超过时不得开始新工作. 各工人间联系充分，保证工作指派不会冲突.

3° 同一作业不因指派多于 1 人而加速，所有工作人员均有相同的能力与效率.

4° 车辆行驶速度规定为 60 公里/小时.

问题分析

问题的关键是将申请修复单位合理排序并合理指派工人. 问题既有离散时间排队系统的特点又有线性规划模型特点. 但在工人的时间约束方面不同于服务系统，在动态性质上又区别于线性规划结构，修复工作的组织应根据三个目标进行：

1° 要考虑受影响地区的社会重要性；

2° 要使修复时间最短，包括完成全部修复所需时间和平均及累计服务时间；

3° 修复费用最小，工作安排应使路途和加班时间最短.

排序方法

选用五种排序方法进行分析比较.

1° 先到先服务(FIFO)，由此可得到各种指标的近似上界.

2° 用 AHP 方法赋权，按权重服务.

采用如下层次结构图(图 9-3).

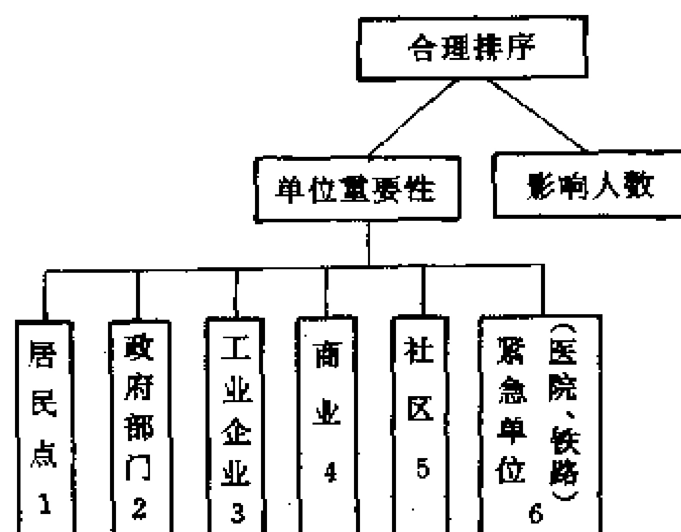


图 9-3 排序层次结构图

表 9-6 合理排序的权重

合理排序	单位重要性	影响人数	优先权重
单位重要性	1	5	5/6
影响人数	1/5	1	1/6

表 9-7 单位重要性的权重

单位重要性	1	2	3	4	5	6	优先权重	准则型权重
1	1	1/6	1/6	1/4	1/3	1/9	0.024843	0.04508
2	6	1	3	5	3	1/7	0.181076	0.32864
3	6	1/3	1	6	4	1/6	0.139315	0.25284
4	4	1/5	1/6	1	3	1/9	0.058630	0.10641
5	3	1/3	1/4	1/3	1	1/9	0.045142	0.08193
6	9	7	6	9	9	1	0.55099	1

对于受影响人数划分为 9 级，对空缺人数用该类型单位的平均人

数替代.

表 9-8　人数转换为 1～9 标度

标度	1	2	3	4	5
人数	10～20	21～50	51～100	101～250	251～500

标度	6	7	8	9
人数	501～800	801～1300	1301～2000	2001～3000

由此得到修复第 j 个单位的优先数(即排序权重)为

$$P_j = TR \times \frac{5}{6} + NP \times \frac{1}{6}$$

其中,TR 为修复单位类型权重,NP 为受影响人数权重.

3° 最短作业时间法(SPT)

该法简单地根据估计作业时间从最短到最长修理时间安排修理工作.

4° 作业松弛法

该法按作业最短松弛时间进行排序,其中作业重要性引用了 AHP 确定的权重 P_i,计算过程如下:

①计算第 i 个单位估计修复时刻 $D_i = E_i/P_i + T_i$. 其中,E_i 为估计修复所需时间,P_i 为作业权重,T_i 为申请报告时间.

②计算松弛时间 $S_i = D_i - E_i$,并根据 S_i 的大小,由小到大进行排序.

5° 比例法

根据比例 $R_i = E_i/P_i$ 由小到大对作业排序,分母中的 P_i 使有较大的权重的单位能排在前面,分子中的 E_i 使需要较短修复时间的单位能较早修复,从而减少系统平均等待时间.

算法描述

在计算机仿真过程中的计算原则是:

1° 初始状态:风暴开始时各中心有一人值班,到清晨 8:00 后每中心有两名工人可指派工作.

2° 一旦申请报告提出，申请单位将按优先权重排到当前等待修复队伍中. 还未参与排队的单位不得进行修复. 当新的一班开始或有工人可指派时，优先权重高的单位首先进行修复，当有多个工人可指派时应使总路途最短. 当某个修理工作不能在一个班内完成时允许超时修理，直到工作完成或超过允许加班时间才返回中心.

3° 对每个工人记下加班时间，对每项作业记下开始修复时间和完成时间以及其他所需信息.

性能测量

修复方案的优劣由下述指标确定.

1° 系统总时间 ST，各项修复工作等待时间之总和，即 $ST=\sum_{i=1}^{N}ST_i$，其中 ST_i＝第 i 个单位修复时刻——报告时间.

2° 系统加权总时间(PST)

$$\mathrm{PST}=\sum_{i=1}^{N}ST_i\times P_i$$

其中 P_i 为 AHP 权重因子.

3° 总修复时间即完成所有修复工作的时间.

4° 费用，即修复期间总工资.

仿真结果如图 9-4 所示.

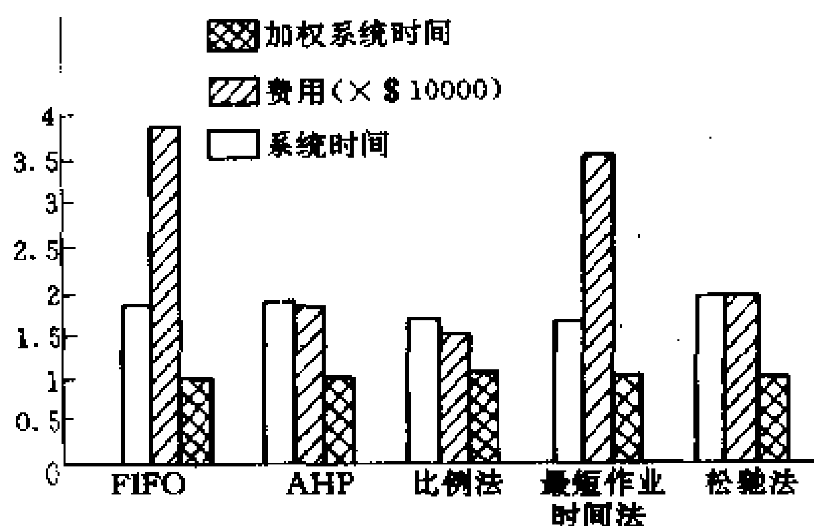


图 9-4　五种方法比较

对五种方法进行比较

1° AHP 法、比例法、作业松弛法由于都考虑了优先权重，故它们的平均加权系统时间较 FIFO 以及最短作业时间法要优.

2° 考虑了估计修复时间进行加权的比例法的平均系统时间较 FIFO 以及 AHP 方法要优.

3° 综合上述两种结果可见，比例法是最优排序方法.

4° 从费用分析看增加最大允许加班时间在支出增加 10%的情况下系统总时间可下降 50%，因此适当加班是可取的.

5° 在这个具有优先权的排队问题中，AHP 中的权重计算可采用归一化的准则型方法，即将最重要单位取权重为 1，然后归一化. 可避免由于类型的增多而改变单位重要性在整个权重中的份额，如表 9-6 最后一列所示.

习　题

1. 某生产线上，零件按泊松分布到达，工序 I 和工序 II 的加工时间分别为正态分布和 β 分布的随机变量，试研究零件入库的分布特征.

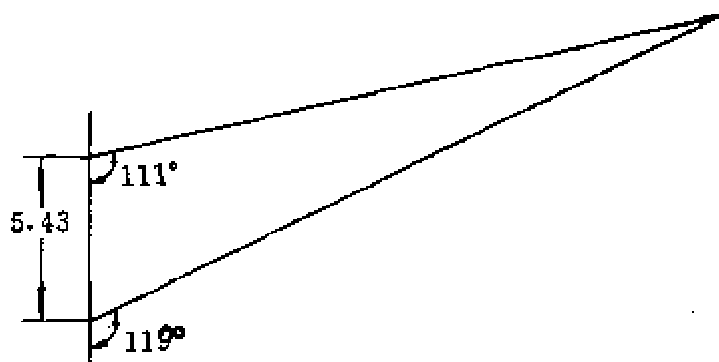


图 9-5

2. 相距 5.43 公里的两个监听站收听到一个短暂的无线电讯号. 收听到讯号的时候测向仪分别定位在 111°和 119°处（如图 9-5），测向仪的精度为 ±2°. 该讯号来自一个毒品交换活跃的地方，据推测该处有一只机动船正等着有人来取毒品. 当时正值黄昏、无风、无潮流. 一架小型直升飞机离开监听站①的简易机场并能准确地沿 111°角方向飞行. 直升飞机的飞行速度是走私

船的三倍.在离船500英尺时船上能听到直升飞机的声音.直升飞机只有一种侦察仪器——探照灯.在200英尺远的地方探照灯只能照明半径为25英尺的圆域.(1)说明飞行员能找到正等着的毒品船的(最小)区域;(2)研究一种直升飞机的最佳搜索方法.在你的计算中要有95%的精度.(MCM1988年A题).

3.煤矿公司经营一个包括一个单个的大型倒煤台在内的装煤设施.当装煤列车到达时,从倒煤台往上装煤.一列标准列车要用3小时装满,而倒煤台的容量是一列半标准列车.每天,铁道部门向这个装煤设施发送i列标准列车.这些列车可在当地时间上午5点到下午8点的任何时间内到达.每列列车有三辆机车.如果一列车到达后因等待装煤而停滞在那里(即处于等待服务状态)的话,铁道部门要征收一种称为滞期费的特别费用,每小时每辆机车5000美元.此外,每周星期四上午11点到下午1点之间有一列大容量列车到达.这种特殊的列车有五辆机车并能装两列标准列车的煤.一个装煤工作班要用6小时直接从煤矿运煤来把空的倒煤台装满.这个工作班(包括它用的设备)的费用是每小时9000美元.可以调用第二个工作班运行一个附加的倒煤台操作系统来提高装煤速度,而费用为每小时12000美元,出于安全的原因,当往倒煤台装煤时,不能往列车上装煤.每当由于往倒煤台装煤而中断往列车上装煤时,就要征收滞期费.

煤矿公司的经理部门要请教你们如何决定该倒煤台的装煤操作的平预期开支,你们的分析应包括以下的问题:

a)应调用几次第二个工作班?

b)预期的月滞期费是多少?

c)如果标准列车能按调度在确切的时间到达,什么样的日调度安排能使装煤费用最少?

d)调用第三个费用为每小时12000美元的倒煤台操作系统工作班,能否降低年操作费用?

e)该倒煤台每天能否再装第四辆标准车的煤?

(MCM1993年B题)

第十章　因子试验法与人工现实法

前面分别讨论了数学建模的三种主要方法:机理分析法、数据分析法和仿真方法. 由于实际问题千差万别,对于某实际问题,能获得的先验知识或试验数据多少不一,因此建模方法也是丰富多样. 本章讨论较复杂系统的数学建模. 一个复杂系统,若能分解成若干子系统,而各子系统具有较充分的先验知识或试验数据,则可先对各子系统建立机理的或统计的模型,然后对复杂系统进行仿真. 在上述要求也达不到的情况下,下面两种方法值得考虑.

一、有计划地作因子试验

当系统现有的数据不能确定个别因素(变量)对系统指标的影响时,这时就有必要在系统上作局部试验(而且可以重复地做),根据试验结果来进行分析求得所需模型结构.

例 10-1　广告费用投资模型

根据一般的想象,总认为广告费用投资越多,销售量就越大,其关系可能是线性的,如图 10-1(a)所示. 运筹学者则不同意此观点,认为它应有饱和段,即应为二次函数,如图 10-1(b)所示. 而心理学家认为应是 Logistci 模型,如图 10-1(c)所示.

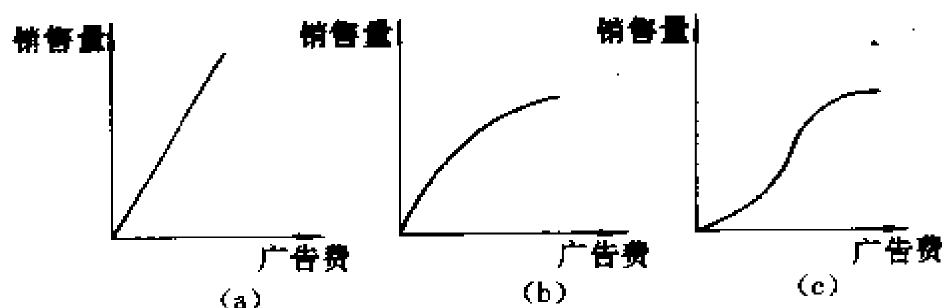


图 10-1　广告投资的假想模型

这三种不同的模型结构需通过试验检验. 选择几个有代表性

的地区对某种商品进行销售试验，结果得到如图 10-2 的曲线.

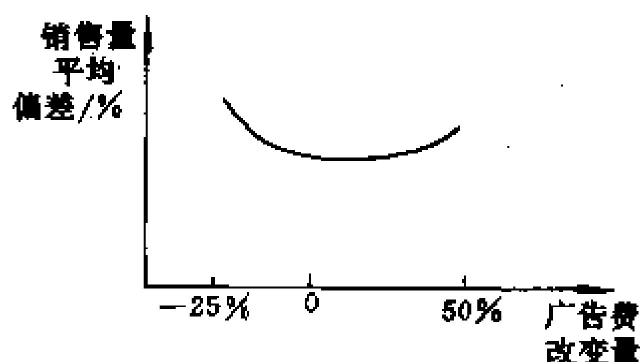


图 10-2 模型的试验

显然试验曲线与假想曲线均不相符. 因此必须进一步进行社会销售行为的分析. 原来社会上有两类人，即较富有的(记为 R)和较穷的(记为 P). 对于较富有者(R)，不论公司是否做广告，还是要购买的，而较穷(P)的由于经济上不富裕，在时间上总有延迟(听取反映和积累钞票). 所以销售曲线应是这两类消费者的迭加，如图 10-3 所示.

进一步的试验可得曲线如图 10-4 所示.

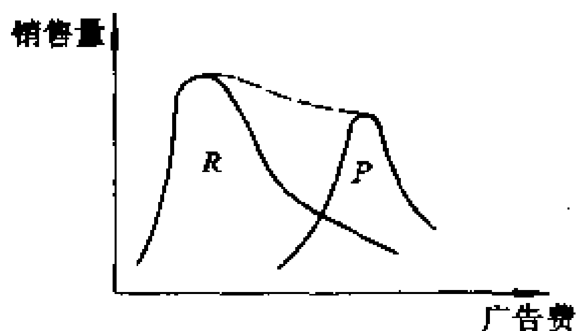


图 10-3 模型的分析

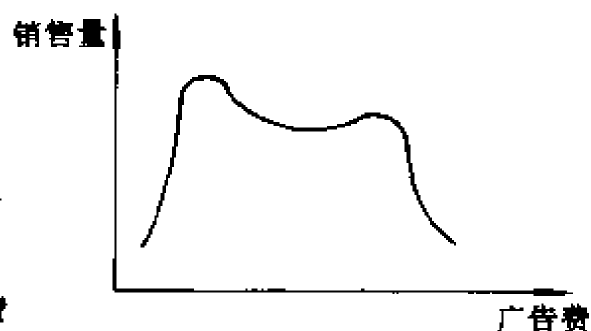


图 10-4 模型的综合

这表明在广告费投资完了之后，在市场上还会维持一段时间的效果，人们对原来广告的印象在广告消除后还会保留一段时间. 根据综合，最后决策应该是波浪式地进行广告投资.

二、利用"人工现实"

在系统结构性质不明确，又无足够的数据可分析，而在系统上

又无法做试验的情况下，可利用“人工现实”着手构造模型.这一方法又称为“人工假设法”或“想定法”(Scenario).

“人工现实”法，是基于对系统过去行为的了解和对未来希望达到的目标，并考虑到系统中有关因素的可能变化，人为地组成一个系统，将“人工现实”适当地分解成一些较为初等的系统(次系统、子系统).这些子系统，或者过去已有研究，或者比较容易形成模型.将系统中的不确定因素假定为若干组确定的值.显然这些人为的假设应该是有根据的合理的.这就形成了一个初步的模型.这样一个人工假定的模型是一个反复修改、不断完善的长期过程.模型的价值取决于对模型中各种因素的数学描述是否正确，模型所使用的数据库是否精确和完善.

人工现实法建模的目的在于为重大决策提供参考，适于一些规模巨大、关系复杂，涉及到人和多方面的不确定因素的系统，如社会经济系统、能源系统、军事系统以及运输贸易系统等等.

参考文献

[1]姜启源. 数学模型. 北京:高等教育出版社,1993. 8.

[2]叶其孝主编. 大学生数学建模竞赛辅导教材. 长沙:湖南教育出版社,1993. 8.

[3]杨启帆,边馥萍. 数学模型. 杭州:浙江大学出版社,1990. 5

[4]Lucas W. F. Models in Applied Mathematics, Vol1-3, Springer-Verlag New York, 1983.

[5]Andress J. G. Mathematical Modelling, Chaped River Press, England, 1976.

[6]Bender E. A. An Introduction of Mathematical Modelling, Wileg Intersience, New York, 1978.

[7]Mischke. Charles R. Mathematical Model Building. The Iowa State University Press, 1980.

[8]江裕钊,辛培清. 数学模型与计算机模拟. 北京:电子科技大学出版社,1989.

[9][美]戴维 C.L.,艾维 E.S. 数学构模原理. 北京:海洋出版社,1985.

[10]苏松基. 系统工程与数学方法. 北京:机械工业出版社,1988.

[11]周仲良,郭镜明译. 美国数学的现在和未来. 上海:复旦大学出版社,1986.

[12]邓越凡译. 数学科学,技术,经济竞争力. 天津:南开大学出版社,1992.

[13]柯朗 R.,希尔伯特 D.. 数学物理方法. 北京:科学出版社,1981.

[14]彼得罗夫斯基 N.Г.. 偏微分方程讲义. 北京:高等教育出版社,1956.

[15]钱颂迪主编. 运筹学. 北京:清华大学出版社,1990.

[16]Qi Huan and Cheng Guohua. The Monte Cario Methed in the Numerical Simulation of the Combustion Proceedings of International Conference on Modelling and Simulation and Control. AMSE 1993.

[17]Sykes. Z. M. On Discrete Stable Population Theory, Biometrics 25, 285—293. 1919.

[18]Kemery, John G, Mathematical Models in the Social Sciences MIT

Press,1978.

[19]齐欢.车用发动机多目标最优控制模型与仿真研究,运筹与决策.成都:成都科技大学出版社,1992.

[20][美]约翰·内特,威廉·沃塞克,迈克尔.H.库特纳.应用线性回归模型.北京:中国统计出版社,1990.

[21]方开泰,全辉,陈庆云.实用回归分析.北京:科学出版社,1988.

[22]韦博成.近代非线性回归分析.上海:东南大学出版社,1989.

[23]黄友谦.曲线曲面的数值表示和逼近.上海:上海科学技术出版社,1984.

[24]李岳生,黄友谦.数值逼近.北京:人民教育出版社,1978.

[25]杨位钦,顾岚.时间序列分析与动态数据建模.北京:北京工业学院出版社,1986.

[26]黄文奇,詹叔浩.求解 Packing 问题的拟物方法.应用数学学报,1979.5.

[27]黄文奇.用于枪炮射表整体解析逼近的一列基函数.应用数学和力学,1981.10.

[28]焦李成.神经网络系统理论.西安:西安电子科技大学出版社,1990.

[29][加拿大]皮洛 E.C.数学生态学.北京:科学出版社,1988.

[30]李楚霖,林少宫.微观经济的数理分析导引.武汉:华中工学院出版社,1985.

[31]Spriet J. A. ,Vansteenkiste G. C. Computer-Aided Modelling and Simulation,Academic Press Inc,1982.

[32]卢开澄.组合数学——算法与分析.北京:清华大学出版社,1983.9.

[33]王莲芬,许树柏.层次分析法引论.北京:中国人民大学出版社,1989.3.

[34]陈秉正.费用分摊问题与群决策方法系统工程理论与实践.1990.10(5).

[35]查有梁,李以渝.数学智慧的横向渗透——数学思想方法论.成都:四川教育出版社,1990.

[36]王仲春,李元中,顾莉蕾,孙名符.数学思维与数学方法论.北京:高等教育出版社,1989.

[37]Foulds L. R. Combinatcorial Optimization for Undergraduates Springer-Verlag,New York,1989.

[38]刘家琦.数学物理方程反问题的分类及不适定问题求解.应用数学与计算数学,1983.4.

[39]张立明.人工神经网络的模型及其应用.上海:复旦大学出版社,1993.

[40]罗发龙，李衍达. 神经网络信号处理. 北京：电子工业出版社，1993.
[41]刘贵忠，邱双亮. 小波分析及其应用. 西安：西安电子科技大学出版社，1992.

后 记

本书介绍了数学建模的主要方法，并采用实例研究法(Case studies)让初学者从中思考体会别人做过的模型，作为当前数学课程的补充，也作为应用数学——应用数学知识解决实际问题——基本训练. 限于篇幅，很多非常有用的模型方法，如物理模型中的量纲分析法、运筹学中的很多有趣模型、动态系统优化中的变分法和最优控制模型等等，都没有在这儿介绍，很多精彩的实例不得不忍痛割爱，所介绍的方法也只涉及到基本概念，并未展开讨论，其中的理论问题则只有放到专门问题中再去研究.

数学建模的理论和方法发展很快，例如小波分析方法和神经网络方法([40]~[42])，它们分别在信号处理、图像处理、量子场论、地震勘探、雷达、天体识别、机器视觉和图像识别、语音识别、自动控制、系统辨识以及决策与评价等众多领域得到广泛的应用. 了解这些新的方法有助于利用数学模型研究解决实际问题.

在计算机飞速发展的今天，数学建模工作不再是一只笔一张纸的纯粹手工劳动，一批计算机辅助建模的软件包成为建模工作者的得力助手，其中在我国使用较为广泛的有 Mathematica™ Math CAD 等，而从决策支持系统角度研究的建模支持系统也开始问世.

数学模型方法，是当代数学领域的一个重要分支，又与各种实际问题涉及的专业知识密不可分，这是数学模型方法与传统的数学课程不同之处，而建模的思维方法和大量技巧也只有在实际建模工作中才能体会到、学习到. 希望这本书能对初学者有所帮助.